

Comité Asesor de Tecnología e Investigación Novedosas y Excepcionales

Informe sobre ciencia de datos
y tecnologías emergentes en la
investigación biomédica

Septiembre de 2023



NATIONAL INSTITUTES OF HEALTH
Comité Asesor de Tecnología e
Investigación Novedosas y Excepcionales
(NExTRAC)



National Institutes
of Health

Lista

Presidencia

BLOSS, Cinnamon, PhD Profesora
Escuela Herbert Wertheim de Salud Pública y Ciencias
de la Longevidad Humana
University of California, San Diego La Jolla, CA 92093

Miembros

BAKKEN, Suzanne, PhD, RN, FAAN, FACMI, FIAHSI
Profesora egresada de la Escuela de Enfermería Profesora de
Informática Biomédica Vagelos College of Physicians &
Surgeons Columbia University
New York, NY 10032

CAINDEC, Karen Presidenta
Junta Directiva
Southcentral Foundation Anchorage, AK 99508

CALLIER, Shawneequa, JD
Profesora Asociada de Investigación Clínica y Liderazgo
Facultad de Medicina y Ciencias de la Salud
George Washington University
Voluntaria especial, National Human Genome Research
Institute
Washington, DC 20037

COLLINS, James, PhD
Profesor de Historia Natural y Medio Ambiente de Virginia
M. Ullman
Jefe de Grupo de la Facultad de Biología del Cambio Global
Escuela de Ciencias de la Vida
Profesor, Centro de Evolución y Facultad Afiliada de
Medicina
Facultad Afiliada al Lincoln Center, Ética Aplicada del Lincoln
Center
Arizona State University
Tempe, AZ 85287

GRONVALL, Gigi Kwik, PhD
Profesora Asociada
Departamento de Salud e Ingeniería Medioambientales
Johns Hopkins Bloomberg School of Public Health
Baltimore, MD 21202

HYUN, Insoo, PhD
Director
Centro de Ciencias de la Vida y Aprendizaje Público
Museum of Science
Boston, MA 02114

KHETERPAL, Sachin, MD, MBA
Decano Asociado de Investigación Informática Profesor
de Anestesiología
University of Michigan Medical School Ann Arbor, MI
48109

LESHNER, Alan I., PhD
Director Ejecutivo, Emérito
American Association for the Advancement of Science
Potomac, MD 20854

OSSORIO, Pilar N., PhD, JD
Profesora de Derecho y Bioética
University of Wisconsin
Madison, WI 53706
Investigador y Responsable del Programa de Ética
Morgridge Institute for Research Madison, WI 53715

OYE, Kenneth, Doctor
Profesor de Ciencias Políticas y Datos, Sistemas y
Sociedad
Director del Programa de Tecnologías Emergentes
Centro de Estudios Internacionales
Massachusetts Institute of Technology
Cambridge, MA 02139

RICHARDSON, Sarah M., PhD
Directora General
Microbyre, Inc.
Berkeley, CA 94710

RILEY, Margaret F., JD
Profesora,
Facultad de Derecho
Profesora de Ciencias de la Salud Pública,
Facultad de Medicina
Profesora de Políticas Públicas,
Escuela Batten de Liderazgo y Políticas Públicas
University of Virginia
Charlottesville, VA 22903

SAUNDERS, Kevin O., PhD
Profesor Asociado
Departamento de Cirugía
Facultad de Medicina de Duke University
Durham, NC 27710

TURNER, Leigh, PhD
Profesor, Programa de Salud, Sociedad y
Comportamiento en Salud Pública
University of California, Irvine
Irvine, CA 92697

WYATT, Letisha R., PhD
Profesora Asociada de Neurología, Facultad de Medicina
Directora de Diversidad en la Investigación,
Investigación e Innovación de la OSHU
Oregon Health & Science University (OHSU)
Portland, Oregon 97239



NATIONAL INSTITUTES OF HEALTH

Comité Asesor de Tecnología e Investigación Novedosas y Excepcionales

Grupo de trabajo sobre ciencia de datos y tecnologías
emergentes en la investigación biomédica



National Institutes
of Health

Lista

Copresidentes

KHETERPAL, Sachin, MD, MBA

Decano Asociado de Investigación Informática Profesor de
Anestesiología

Facultad de Medicina de University of Michigan

Ann Arbor, MI 48109

OSSORIO, Pilar N., PhD, JD

Profesora de Derecho y Bioética University of Wisconsin

Madison, WI 53706

Investigador y Director del Programa de Ética de Lead

Morgridge Institute for Research

Madison, WI 53715

Miembros

ALBRITTON, Lorraine M., PhD*

Profesora Emérita

Departamento de Microbiología, Inmunología y Bioquímica

Facultad de Medicina

The University of Tennessee Health Science Center

Memphis, TN 38163

BAKKEN, Suzanne, PhD, RN, FAAN, FACMI, FIAHSI

Profesora Egresada de la Escuela de Enfermería

Profesora de Informática Biomédica

Vagelos College of Physicians & Surgeons

Universidad de Columbia

New York, NY 10032

BLOSS, Cinnamon, PhD

Profesora

Escuela Herbert Wertheim de Salud Pública y

Ciencias de la Longevidad Humana

University of California, San Diego

La Jolla, CA 92093

LESHNER, Alan I., PhD

Director Ejecutivo, Emérito

American Association for the Advancement of Science

Potomac, MD 20854

MANJI, Hussein K., MD, FRCPC*

Profesor Visitante,

Oxford University

Copresidencia,

Misión de Salud Mental del Reino Unido

Exjefe Terapéutico Global de Neurociencia

Janssen Research & Development

Oxford, Inglaterra

MATHEWS, Debra J. H., PhD, MA*

Directora Asociada de Investigación y Programas

Berman Institute of Bioethics

Profesora Asociada de Medicina Genética

Johns Hopkins University

Baltimore MD 21205

RILEY, Margaret F., JD

Profesora,

Facultad de Derecho

Profesora de Ciencias de la Salud Pública,

Facultad de Medicina

Profesora de Políticas Públicas

Escuela Batten de Liderazgo y Políticas Públicas

University of Virginia

Charlottesville, VA 22903

*Miembros *Ad Hoc*

CONTENIDO

Resumen ejecutivo.....	1
I. Introducción	4
II. Antecedentes y visión general	7
III. Preguntas de investigación sobre ciencia de datos	9
A. Introducción y antecedentes	9
B. Tipos de preguntas de investigación a considerar a través de la participación pública	11
IV. Participación pública para informar sobre las implicaciones de las preguntas de investigación y las tecnologías	15
A. Recogida de información sobre la participación pública y métodos de análisis	15
B. Conversaciones comunitarias en Estados Unidos.....	16
C. Taller: utilizar la participación pública para informar sobre el uso de datos en la investigación biomédica	22
V. Observaciones generales, conclusiones y recomendaciones para los NIH.....	26
VI. Limitaciones de este esfuerzo y posibles orientaciones futuras.....	32
VII. Conclusión.....	33
VIII. Agradecimientos.....	35
IX. Anexo 1: descripción detallada de la Fase 2.....	36
A. Calendario de conversaciones comunitarias y seminarios web.....	36
B. Estudios de casos y análisis de las conversaciones mantenidas en la fase 2	37
X. Apéndice 2: la IA generativa como caso práctico para las recomendaciones de los WG	42
A. Introducción.....	42
B. Caso práctico	42

Resumen ejecutivo

Los avances en la ciencia de datos son muy prometedores para mejorar nuestra comprensión de la salud y las enfermedades humanas, pero ciertas aplicaciones de la ciencia de datos también pueden socavar la autonomía, la privacidad o la experiencia de respeto de los participantes en la investigación cuando contribuyen a la investigación biomédica.¹ Utilizadas de forma inadecuada, las herramientas desarrolladas mediante la ciencia de datos emergente podrían afianzar o exacerbar las desigualdades en materia de salud o atención médica. Los Institutos Nacionales de Salud (National Institutes of Health, NIH) tratan de comprender las implicaciones políticas y éticas de las tecnologías emergentes de la ciencia de datos, y cómo las tecnologías emergentes permiten el uso de conjuntos de datos de fuentes dispares (p. ej., datos procedentes de entornos de investigación y de no investigación) para la investigación en salud. Los NIH también pretenden anticipar los posibles beneficios y riesgos de la ciencia de datos emergente para los participantes en la investigación, las familias, las poblaciones, los investigadores y la sociedad. Para ayudar a evaluar los recientes avances de la ciencia de datos en la investigación biomédica, los NIH recurrieron al Comité Asesor de Tecnología e Investigación Novedosas y Excepcionales (Novel and Exceptional Technology and Research Advisory Committee, NExTRAC), un comité asesor federal que ofrece orientación al director de los NIH sobre el uso de tecnologías emergentes en la investigación biomédica, para ayudar a garantizar que la investigación progrese de forma responsable y cumpla las expectativas públicas de la inversión de la Nación en investigación. El NExTRAC se encargó de lo siguiente:

- Definir y caracterizar los tipos de preguntas de investigación que requieren una granularidad y agregación crecientes de datos sobre personas que probablemente se aborden mediante tecnologías emergentes.
- Para las preguntas y tecnologías definidas anteriormente, consultar con las partes interesadas para debatir y evaluar el valor y las posibles implicaciones para las personas, los grupos y la sociedad.

Para ello, el NExTRAC creó el Grupo de Trabajo (Working Group, WG) sobre Ciencia de Datos y Tecnologías Emergentes. El WG dividió su esfuerzo en dos fases separadas. Como parte de la Fase 1, el WG se reunió con expertos en ciencias de datos relacionados con la salud para analizar los esfuerzos de investigación actuales y solicitar aportes sobre tecnologías emergentes y asuntos éticos en este campo. A partir de estos debates y de un análisis más detallado, el Comité identificó tres áreas temáticas de investigación generales y los correspondientes tipos de preguntas de investigación, que se exponen en la siguiente tabla:

Tabla 1: Tipos de preguntas de investigación definidas en la Fase 1	
Áreas temáticas de investigación	Tipos de preguntas de investigación
Datos: uso de datos novedosos ajenos al sistema de atención médica tradicional.	¿Cómo se utilizan los datos personales ² recogidos fuera del sistema de salud tradicional o del entorno de investigación (p. ej., a través de rastreadores de actividad física, aplicaciones y publicaciones en redes sociales) para estudiar cuestiones relacionadas con la salud y predecir riesgos de salud individuales, familiares, de grupo o de salud pública?
	¿Cómo pueden/deben utilizarse otros datos sobre consumidores y estilos de vida procedentes de fuentes no específicamente de salud (p. ej., sensores en el hogar y datos sobre tarjetas de crédito y recompensas a consumidores) para estudiar cuestiones relacionadas con la salud y predecir riesgos de salud?
	¿Cómo pueden/deben integrarse los datos de salud con los datos sobre los determinantes sociales de la salud (Social Determinants Of Health, SDOH) para permitir una mejor predicción del riesgo y el desarrollo de algoritmos predictivos?
Algoritmos:	¿Cuál es el papel de las tecnologías algorítmicas basadas en la informática, como la inteligencia artificial (IA), el procesamiento del lenguaje natural (natural language processing, NLP) (incluidos los grandes modelos lingüísticos como ChatGPT), aprendizaje

¹ La investigación biomédica es “la investigación que aborda los mecanismos que subyacen a la formación y función de los organismos vivos, desde el estudio de moléculas individuales hasta las complejas funciones integradas de los seres humanos, y contribuye profundamente a nuestro conocimiento de cómo las enfermedades, los traumas o los defectos genéticos alteran los procesos fisiológicos y de comportamiento normales”. National Research Council (2005). *Advancing the Nation's Health Needs: NIH Research Training Programs*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/11275>.

² En este informe, el Comité considera “datos de salud personales” los datos recogidos sobre personas o grupos y analizados con el fin de abordar cuestiones relacionadas con la salud en la investigación biomédica. Esta amplia definición engloba los datos que se recopilan en clínicas y entornos de atención médica (el sistema de salud “tradicional”) y los datos de salud generados por la persona (es decir, recopilados por la persona a través de dispositivos personales diseñados para recoger métricas relacionadas con la salud). Puede tratarse de observaciones, opiniones, percepciones, motivaciones, comportamientos, etc.

Uso de modelos y algoritmos	automático (machine learning, ML) y el análisis automatizado de imágenes en la investigación biomédica y la toma de decisiones de salud?
	¿Cómo puede/debe desplegarse la NLP para analizar los datos de los sistemas de salud (p. ej., historias clínicas electrónicas [electronic health records, EHR], datos de seguros médicos y datos de farmacias) con el fin de obtener información sobre los síntomas de los pacientes y la clasificación de las enfermedades?
Integración: vinculación y agregación de conjuntos de datos dispares procedentes de varias fuentes.	¿Existen oportunidades para estandarizar los formatos de los datos —o desplegar tecnologías de estandarización— de modo que los datos de diferentes países y sistemas de atención médica puedan agregarse, vincularse y compartirse entre poblaciones?
	¿Qué conjuntos de datos dispares (y potencialmente contradictorios) (p. ej., genómica, proteómica, información e imágenes clínicas) pueden vincularse y combinarse mediante agregadores [automatizados] de datos?
	¿Cómo pueden/deben utilizarse las bibliotecas de salud personales para combinar la información de salud de las personas a través de múltiples flujos de datos diferentes con el fin de informar sobre los resultados de salud?
	¿Cómo puede/debe utilizarse la vinculación de registros para preservar la privacidad (Privacy-Preserving Record Linkage, PPR) para combinar datos sobre personas procedentes de múltiples fuentes y con diferentes identificadores para la medicina de precisión y la salud pública?
	¿Cómo la vinculación y agregación de datos pueden/deben afectar el contexto de la investigación (p. ej., clínico y de salud pública) y el estado de consentimiento de los participantes?

Los debates adicionales sobre la información recopilada durante la Fase 1 condujeron a las conclusiones que figuran en la tabla 2.

En la Fase 2, para informar los debates del NEXTRAC y el WG, los NIH convocaron primero una serie de conversaciones comunitarias con miembros del público para recabar opiniones sobre las preguntas de investigación de la Fase 1 y su posible impacto en las personas y las comunidades. El WG pidió deliberadamente a los NIH que incluyeran en estos compromisos a miembros de comunidades históricamente marginadas en la investigación biomédica. El objetivo de estos encuentros era reunirse con el público en sus propias comunidades, entablar un diálogo respetuoso con los participantes, sentar las bases de relaciones duraderas y fomentar la confianza del público en la investigación biomédica.

En las conversaciones con la comunidad surgieron tres temas principales. En primer lugar, el apoyo de los miembros de la comunidad a las iniciativas de intercambio y vinculación de datos dependía de quién recibiera los datos, de si los receptores se consideraban dignos de confianza y de si habría un retorno significativo de valor para las personas y las comunidades. En segundo lugar, muchos miembros de la comunidad expresaron su deseo de que se introdujeran cambios en el proceso de consentimiento informado para que la gente supiera más sobre cómo y cuándo se utilizan sus datos (tanto si se recogen inicialmente para investigación como para otros fines). Además, algunas personas querían un mayor control sobre el uso de sus datos personales. En tercer lugar, aunque muchos miembros de la comunidad expresaron cierto entusiasmo por los nuevos usos de los datos en las tres áreas temáticas de investigación, muchos también solicitaron una mayor transparencia sobre el uso de los datos de investigación, garantías de salvaguardias y beneficios adicionales para la investigación que hicieran que la participación mereciera los riesgos. A través de las deliberaciones sobre los comentarios aquí señalados, el NEXTRAC informa las siguientes conclusiones (tabla 2).

Fase de esfuerzo	Conclusiones
Fase 1 Debates de los grupos de trabajo con expertos en la materia	1. Las tecnologías emergentes para capturar, almacenar y analizar datos de salud personales (p. ej., tecnologías portátiles, aplicaciones e inteligencia artificial) ofrecen abundantes recursos para la investigación. Sin embargo, se necesita más trabajo político y ético para aprovechar responsablemente todas sus ventajas.
	2. Los nuevos métodos para armonizar y agregar datos procedentes de distintos estudios y de todo el mundo son muy prometedores para la medicina de precisión y la salud pública. Aun así, es necesario introducir mejoras para limitar la variación que puede reducir la significación, la precisión de los resultados y el carácter general de las conclusiones.
Fase 2 Participación pública	3. Existe un desajuste entre las expectativas de algunos miembros del público y algunas políticas y programas de los NIH sobre el uso y análisis de datos personales de salud para la investigación biomédica.

con miembros de la comunidad y académicos	4. Es importante consultar a los participantes en la investigación, a las personas cuyos datos se utilizan en la investigación secundaria, a las familias y a las comunidades a lo largo de todo el proceso de investigación biomédica cuando se trata de nuevos tipos, usos y análisis de datos.
	5. Se necesita más transparencia en torno al proceso de investigación, cómo se recogerán, analizarán y utilizarán los datos, y los riesgos y beneficios para los participantes en la investigación.
	6. La normativa y los procedimientos actuales para el uso de datos de investigación biomédica ofrecen vías para que los datos se reutilicen en la investigación sin consentimiento ni supervisión adecuada.
	7. La vinculación de datos en ausencia de un consentimiento informado sólido suscita preocupación por la privacidad de los datos, la transparencia y el respeto a los participantes en la investigación.
	8. Más allá de las promesas potenciales de la medicina de precisión, los participantes en la investigación merecen un mayor retorno de valor para ellos y sus comunidades.
	9. Proporcionar a los participantes en la investigación beneficios económicos o un mayor control sobre los datos personales puede reforzar el apoyo de la comunidad a la investigación biomédica realizada por entidades con ánimo de lucro o en asociaciones público-privadas.

A la luz de estos resultados, el Comité presenta ocho recomendaciones de alto nivel centradas en los siguientes temas generales: **fiabilidad y transparencia en la investigación, respeto por los participantes y las comunidades, mejora del consentimiento informado y del control del intercambio y la vinculación de datos, y retribución de la participación en la investigación, el intercambio y la vinculación de datos** (Tabla 3).

Tabla 3: recomendaciones basadas en las conclusiones
1. Los NIH deberían catalogar, evaluar y, en su caso, coordinar los esfuerzos entre las iniciativas actuales que abordan el compromiso público, las cuestiones éticas y tecnológicas en la ciencia de datos.
2. Cuando los NIH invierten en la participación pública, especialmente en torno a las tecnologías emergentes, la agencia debe estar preparada para actualizar los procesos y la infraestructura y para revisar sus programas y políticas de acuerdo con las expectativas públicas y los principios de justicia social, o explicar claramente al público por qué la agencia no puede revisar su curso.
3. Los NIH, en consulta con otras agencias federales según proceda, deberían alinear sus políticas sobre datos y gobernanza de la investigación con valores públicos ampliamente extendidos.
4. Los NIH deben ayudar a establecer relaciones a largo plazo entre los miembros del público y las comunidades de investigación para facilitar la participación pública sostenida en las cuestiones en evolución relativas al uso, análisis, gestión y reutilización de los datos personales de salud en la investigación biomédica.
5. Los NIH deberían animar a los investigadores a abordar de forma prospectiva en sus programas y propuestas de investigación las consideraciones bioéticas en áreas emergentes de la ciencia de datos que son particularmente importantes para las comunidades participantes, tal y como se refleja en el compromiso público.
6. Los NIH, en consulta con otras agencias federales según proceda, deben desarrollar y difundir las mejores prácticas para informar a las comunidades participantes sobre cómo se gestionan, reutilizan y vinculan sus datos personales de salud con otros datos, especialmente cuando estas mejores prácticas van más allá de los requisitos reglamentarios.
7. Los NIH deben aprovechar la investigación previa sobre el retorno de los resultados y apoyar una mayor participación pública para informar sobre las mejores prácticas para el retorno de los resultados de la investigación biomédica utilizando la ciencia de datos emergente y para determinar las necesidades de infraestructura de implementación.
8. Los NIH deben apoyar la exploración de métodos para proporcionar beneficios y retorno de valor más allá de la devolución de los resultados de la investigación a los participantes y sus comunidades, cuando sea apropiado y coherente con la normativa vigente.

El enfoque, las observaciones y el razonamiento del Comité para llegar a sus conclusiones y recomendaciones se describen en su totalidad a continuación, con información contextual en los apéndices.

I. Introducción

Mejorar nuestra comprensión de la salud y las enfermedades humanas exige acceder a los datos y utilizarlos plenamente. Cada vez más, los investigadores buscan estos datos en lugares ajenos al sistema de atención médica tradicional, como a través de aplicaciones de salud,³ tecnologías portátiles,⁴ redes sociales⁵ y otras nuevas fuentes de datos. Muchos esperan que estos conjuntos de datos de mayor dimensión ofrezcan una imagen más completa de la salud y el bienestar de una persona y estimulen avances en la medicina personalizada que puedan adaptar los tratamientos a las necesidades de cada persona. Los nuevos métodos de IA⁶ y ML⁷ permiten a los investigadores manejar estos conjuntos de datos y prometen nuevos conocimientos para la investigación y decisiones más rápidas en la clínica. Los investigadores también han desarrollado nuevas formas de vincular y agregar datos personales y hacerlos más útiles, entre otras cosas para hacer frente a emergencias de salud pública como la reciente pandemia de COVID-19.⁸

Aunque estos avances pueden mejorar la investigación biomédica, plantean cuestiones difíciles a los responsables políticos y a los ciudadanos. ¿Qué piensa la gente sobre los riesgos y beneficios de compartir su información de salud personal en estos nuevos contextos? ¿Cómo pueden afectar estos nuevos usos de los datos a las personas y las comunidades, y cómo deben ajustarse en consecuencia la recopilación de datos y las políticas en torno a su uso? ¿Existe una protección adecuada de la privacidad? ¿Cómo deben cambiar las políticas para seguir el ritmo de la ciencia y la tecnología emergentes?

Para ayudar a abordar cuestiones como estas, los NIH recurren a menudo a comités consultivos como el NExTRAC. El objetivo del NExTRAC, tal como se indica en los estatutos del Comité, es,

“...asesorar al Director [de los NIH] en asuntos relacionados con la realización y supervisión de investigaciones que impliquen tecnologías emergentes en la ciencia biomédica (también denominadas biotecnologías emergentes). El Comité abordará las cuestiones científicas, de seguridad, éticas y sociales asociadas a las áreas de investigación biotecnológica emergente para las que los NIH soliciten asesoramiento u orientación.”

Aquí, los NIH buscan anticipar cómo las tecnologías emergentes pueden permitir la combinación y el uso de conjuntos de datos humanos, en particular de fuentes dispares (p. ej., entornos de investigación y no de investigación), para anticipar los beneficios y riesgos potenciales para los participantes en la investigación, las familias, las poblaciones y la sociedad. Los NIH tendrán en cuenta esta información a medida que la agencia desarrolle e implemente estrategias de investigación y marcos políticos, informados por las aportaciones del público, para garantizar que la ciencia avanza de forma responsable y cumple con las expectativas de la inversión de la Nación en investigación. En julio de 2021, el NExTRAC se encargó de lo siguiente:

³ Chaturvedi, R. R., Angrisani, M., Troxel, W. M., Gutsche, T., Ortega, E., Jain, M., Boch, A. y Kapteyn, C. (2023). *American Life in Realtime: a benchmark registry of health data for equitable precision health*. *Nature Medicine*, 29, 283-286. <https://doi.org/10.1038/s41591-022-02171-w>; Stern, A. D., Brönneke, J., Debatin, J. F., Hagen, J., Matthies, H., Patel, S., Clay, I., Eskofier, B., Herr, A., Hoeller, K., Jaksa, A., Kramer, D. B., Kuhlstedt, M., Lofgren, K. T., Mahendraratnam, N., Muehlan, H., Reif, S., Riedemann L. y Goldsack, J. C. (2022). *Advancing digital health applications: priorities for innovation in real-world evidence generation*. *Lancet Digital Health*, 4, e200-e206. [https://doi.org/10.1016/S2589-7500\(21\)00292-2](https://doi.org/10.1016/S2589-7500(21)00292-2).

⁴ Abramson, A., Chan, C. T., Khan, Y., Mermin-Bunnell, A., Matsuhisa, N., Fong, R., Shad, R., Hiesinger, W., Mallick, P., Gambhir, S. S. y Bao, Z. (2023). *A flexible electronic strain sensor for the real-time monitoring of tumor regression*. *Science Advances*, 8, eabn6550. <https://doi.org/10.1126/sciadv.abn6550>; Mannhart, D., Lischer, M., Knecht, S., du Fay de Lavallaz, J., Strebel, I., Serban, T., Vögeli, D., Schaer, B., Osswald, S., Mueller, C., Kühne, M., Sticherling, C. y Badertscher, P. (2023). *Clinical Validation of 5 direct-to-consumer wearable smart devices to detect atrial fibrillation: BASEL Wearable Study*. *JACC Clinical Electrophysiology*, 9, 232-242. <https://doi.org/10.1016/j.jacep.2022.09.011>.

⁵ Klein, A. Z., Magge, A., O'Connor, K. y Gonzalez-Hernandez, G. (2022). *Automatically identifying Twitter users for interventions to support dementia family caregivers: Annotated data set and benchmark classification models*. *JMIR Aging*, 5, e39547. <https://doi.org/10.2196/39547>; Nebeker, C., Dunseath, S. E. y Linares-Orozco, R. (2020). *A retrospective analysis of NIH-funded digital health research using social media platforms*. *Digital Health*, 6, 2055207619901085. <https://doi.org/10.1177/2055207619901085>.

⁶ U.S. Government Accountability Office and National Academy of Medicine (2022). *Artificial Intelligence in Health Care: Benefits and Challenges of Machine Learning Technologies for Medical Diagnostics*. <https://www.gao.gov/products/gao-22-104629>.

⁷ Nestor, B., Hunter, J., Kainkaryam, R., Drysdale, E., Inglis, J. B., Shapiro, A., Nagaraj, S., Ghassemi, M., Foschini, L. y Goldenberg, A. (2023). *Machine learning COVID-19 detection from wearables*. *Lancet Digital health*, 5, e182-e184. [https://doi.org/10.1016/S2589-7500\(23\)00045-6](https://doi.org/10.1016/S2589-7500(23)00045-6).

⁸ National Academy of Medicine (2022). *Sharing Health Data: The Why, the Will, and the Way Forward*. Washington, DC: National Academy of Sciences. <https://nam.edu/wp-content/uploads/2022/01/Sharing-Health-Data-prepub-FINAL.pdf>.

- Definir y caracterizar los tipos de preguntas de investigación que requieren una granularidad y agregación crecientes de datos sobre personas que probablemente se aborden mediante tecnologías emergentes, teniendo en cuenta:
 - Objetivos de estos estudios de investigación y cómo contribuyen a la misión de los NIH
 - Tecnologías emergentes que pueden generar conjuntos de datos potencialmente sensibles
 - Tipos de datos generados y sus fuentes (p. ej., dispositivos de salud digitales y plataformas de EHR), haciendo énfasis en la exploración de nuevos tipos de datos o fuentes únicas
 - Plataformas y herramientas de ciencia de datos que facilitan el acceso, la combinación y el análisis de datos (p. ej., IA y computación en nube)
- Para las preguntas y tecnologías definidas anteriormente, consulte con las partes interesadas para debatir y evaluar el valor y las posibles implicaciones para las personas, los grupos y la sociedad, teniendo en cuenta:
 - Actitudes y perspectivas sobre el intercambio de datos de los participantes para avanzar en la investigación biomédica, especialmente desde el punto de vista del equilibrio entre los riesgos de la investigación (p. ej., la privacidad y la autonomía) y los resultados de la investigación
 - Cómo pueden evolucionar estas perspectivas en función del contexto de quién debe beneficiarse o asumir el riesgo, ya sea a nivel individual, a través de la comunidad o de las expectativas de la sociedad en general sobre el avance de la salud pública

Para ello, se pidió al Comité que consultara con miembros del público, incluidos, entre otros, participantes en investigaciones, grupos de pacientes, expertos en ética y privacidad, científicos de datos, desarrolladores de tecnología de distintos sectores y funcionarios de salud pública.

Como se explica en la Sección II, se creó un WG NExTRAC para llevar a cabo el encargo en dos fases distintas: en primer lugar, identificar tipos de preguntas de investigación (Sección III) y, en segundo lugar, debatir esas preguntas de investigación con miembros del público (Sección IV; figura 1). Para abordar la segunda parte del encargo, el WG incluyó intencionadamente voces que suelen estar infrarrepresentadas en la investigación biomédica y escuchó a miembros del público en conversaciones facilitadas en sus propias comunidades. Además, en apoyo del esfuerzo, los NIH organizaron un taller para escuchar a académicos, desarrolladores y médicos-científicos, centrando al mismo tiempo los comentarios de la comunidad. El enfoque de la participación pública se explica con más detalle en la Sección IV. El Comité se basó en gran medida en los comentarios de estos debates para generar recomendaciones que los NIH debían tener en cuenta (Sección V). El objetivo de estas recomendaciones es facilitar el desarrollo y la aplicación de políticas, programas y procesos que ayuden a los investigadores a integrar las perspectivas públicas y las implicaciones éticas, jurídicas y sociales de los nuevos usos y análisis de datos en la investigación biomédica.⁹

⁹ Cabe destacar que, si bien se habló de NLP y de grandes modelos lingüísticos, el debate sobre el uso de la IA generativa en la investigación biomédica y la sociedad siguió siendo limitado. ChatGPT (y otras tecnologías similares de inteligencia artificial) cobraron protagonismo cuando el WG estaba terminando su tarea. El WG ha utilizado programas como ChatGPT como Caso práctico para sus recomendaciones a los NIH, tal y como se describe en el apéndice 2.

Flujo de trabajo y resultados de NExTRAC sobre ciencia de datos y tecnologías emergentes

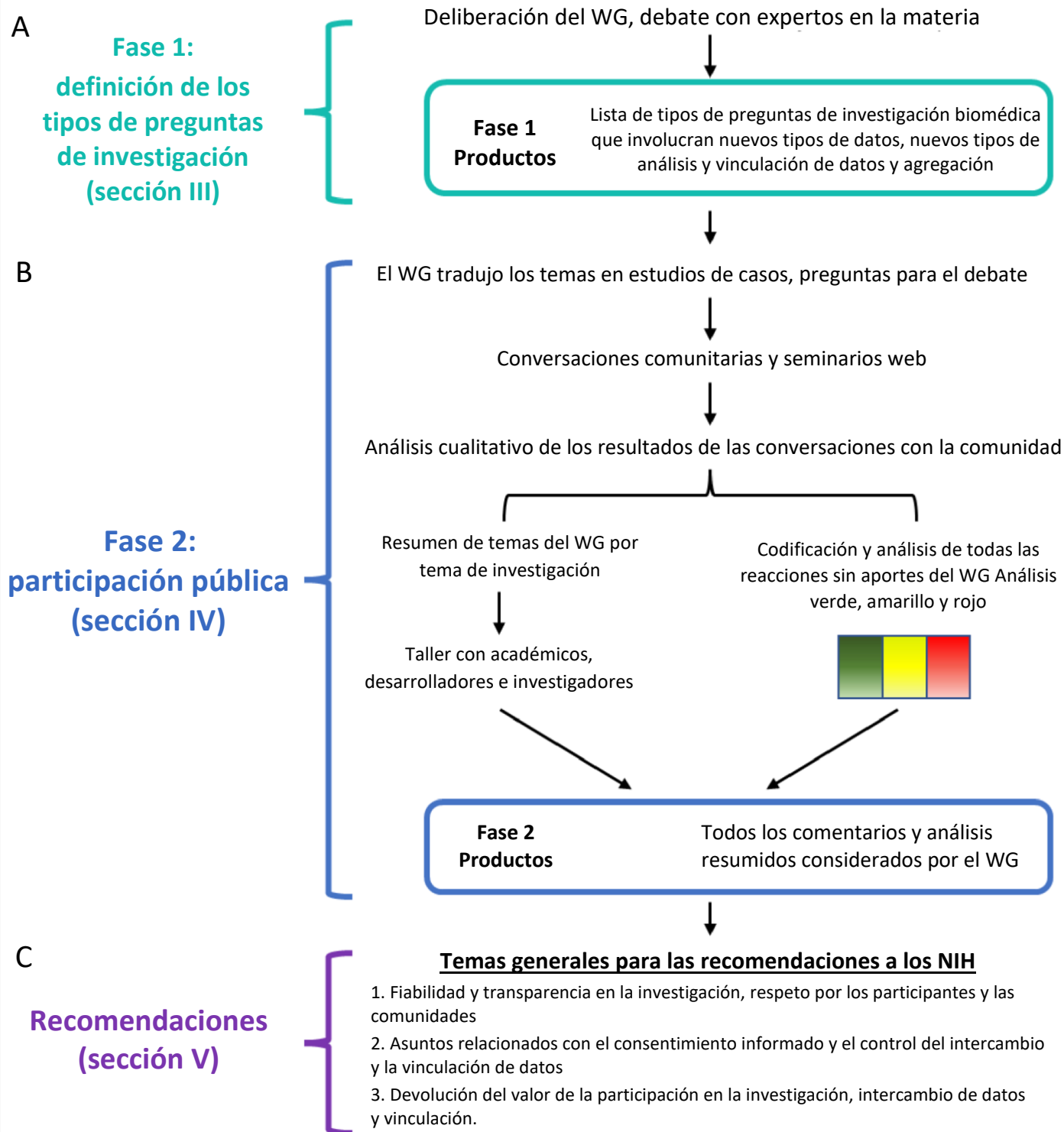


Figura 1: flujo de trabajo y resultados. A) En la Fase 1, el WG definió los temas de investigación y los tipos de preguntas de investigación basándose en deliberaciones internas y debates con expertos en la materia. B) En la fase 2, el WG recopiló información de las conversaciones con la comunidad y de un taller con académicos, que fue resumida por el WG y codificada y analizada sin la aportación del WG. C) Se elaboraron recomendaciones para los NIH sobre tres temas generales.

II. Antecedentes y visión general

El marco para las actividades del WG lo proporcionó el responsable del NExTRAC, así como un informe anterior del Comité, el *Informe para establecer un marco NExTRAC*.¹⁰ El cargo se dividió en dos partes, que sirvieron para organizar las actividades del WG. Se encargó al WG que, en primer lugar, identificara los tipos de preguntas de investigación y, en segundo lugar, que debatiera esas preguntas de investigación con los miembros pertinentes del público. Como se indica en la introducción, el WG ha llevado a cabo estas dos tareas en fases separadas (Fase 1 y Fase 2; figura 1). Para orientar las deliberaciones, el WG tomó nota de varios temas posibles que ilustran las cuestiones esenciales para estructurar sus deliberaciones (tabla 4). A partir de estos temas, el WG identificó posteriormente los tipos de preguntas de investigación (Fase 1; Sección III) y elaboró materiales para evaluar las actitudes del público (Fase 2; Sección IV; apéndice 1).

Tabla 4. Temas para orientar las deliberaciones	
Temas	Pregunta evaluativa
Novedad	¿Las tecnologías o los datos existen desde hace mucho tiempo?
Contexto	¿Se recopilan, utilizan o aplican los datos en nuevos ámbitos?
Complejidad	¿Los datos son especialmente difíciles de analizar, incluso en pequeñas cantidades?
Sensibilidad	¿Los datos revelan información personal sobre personas, familias o grupos?
Justicia	¿Existe la posibilidad de que los datos produzcan una distribución injusta de riesgos y beneficios?
Acceso	¿Hay falta de acceso a los datos, a las tecnologías subyacentes o a los beneficios derivados de su uso?
Fiabilidad	¿Los datos son predictivos y útiles para abordar las preguntas de investigación pertinentes? ¿Se puede confiar en los datos para tomar decisiones?
Consecuencias imprevistas	¿Los datos tienen un potencial significativo de ser utilizados de forma distinta a la prevista originalmente y de causar daños por esos usos?
Control	¿Las personas pueden ejercer control sobre los usos posteriores de los datos?
Agregación	¿Se combinan los datos con otros tipos de conjuntos de datos o se vinculan fácilmente a ellos?
Identificabilidad	¿Los datos contienen información identificable, por sí solos o combinados con otra información fácilmente disponible?
Privacidad, confidencialidad	¿Los datos son vulnerables al acceso no autorizado, la divulgación y la revelación de información personal?

El WG se centró en tres áreas temáticas de investigación para informar sobre cuestiones relacionadas con la salud:

- **Datos:** uso de datos novedosos ajenos al sistema de atención médica tradicional.
- **Algoritmos:** Uso de modelos y algoritmos
- **Integración:** vinculación y agregación de conjuntos de datos dispares procedentes de varias fuentes.

Este marco de Datos, Algoritmos e Integración pretendía abarcar todo el ciclo de vida de los datos, desde su recogida hasta su almacenamiento, análisis, puesta en común y usos posteriores. El primer componente, los datos novedosos, aborda las nuevas fuentes de información que se utilizan en la investigación biomédica y conductual. El segundo componente, modelos y algoritmos, se pregunta por las nuevas formas en que se analizan los datos de la investigación biomédica, tanto los nuevos como los históricos, y en algunos casos se incorporan a la atención médica. El tercer componente, la integración, abarca las nuevas formas en que los conjuntos de datos se combinan, enlazan y agregan para dar lugar a nuevos descubrimientos. Estos temas de investigación se tratan con más detalle en la Sección III.

El WG también trató de evitar la duplicación de esfuerzos anteriores y actuales en todo el gobierno de Estados Unidos. Por ejemplo, se dejó de lado la investigación centrada específicamente en los datos genómicos, dado que hay muchas otras iniciativas en curso en este ámbito. Además, el WG observó que en los NIH y en todo el gobierno de Estados Unidos había varias iniciativas centradas en la IA y quiso asegurarse de que ninguna recomendación sobre IA se limitara

¹⁰ Novel and Exceptional Technology and Research Advisory Committee (diciembre de 2020). *Report to Establish a NExTRAC Framework*. https://osp.od.nih.gov/wp-content/uploads/NExTRAC-Framework-Report_FINAL_508.pdf.

a repetir las conclusiones de informes anteriores. Por último, el WG hizo menos énfasis en los asuntos planteados por el intercambio internacional de datos.

Dado el importante papel que desempeñan las normativas de Estados Unidos (p. ej., la Ley de Transferencia y Responsabilidad de Seguros Médicos [Health Insurance Portability and Accountability Act, HIPAA] y la Regla Común) a la hora de determinar cómo se utilizan los datos en la investigación financiada por los NIH, el WG consideró en cambio el intercambio de datos en un contexto nacional.

A. Enfoque para la Fase 1

Como se indica en el *Informe para establecer un marco NExTRAC*, el Comité utiliza la exploración del horizonte para identificar las tecnologías emergentes que se implantarán en los próximos 5-10 años. Como punto de partida, el WG celebró reuniones de intercambio de ideas para compartir perspectivas sobre las cuestiones más urgentes de la ciencia de datos en la investigación biomédica e identificar lagunas en los conocimientos del WG. Para colmar estas lagunas, el WG realizó una revisión bibliográfica e identificó a expertos en temas relacionados con el cometido del NExTRAC. A lo largo de la Fase 1, el WG examinó los siguientes temas:¹¹

- Uso de las redes sociales y otros datos generados por los usuarios para el diagnóstico médico y la vigilancia de la salud pública
- Uso de la IA y el ML en la toma de decisiones médicas
- Plataformas de datos de salud e intercambio de datos federados¹²
- Datos sobre consumidores y estilos de vida procedentes de fuentes no específicas de la salud
- Nuevos flujos de datos gracias a los dispositivos portátiles, los teléfonos inteligentes e internet.

En julio de 2022, el WG presentó al NExTRAC un borrador de lista de tipos de preguntas de investigación, basado en su análisis articulado en la tabla 1 (Resumen Ejecutivo) y en un resumen con viñetas (Sección III). Como ya se ha indicado, el WG dio prioridad a un subconjunto de estas preguntas de investigación y las tradujo en estudios de casos para la participación pública en la Fase 2.

B. Enfoque para la Fase 2

En la Fase 2 del proyecto, el WG solicitó la opinión de diversos sectores de la población y de las comunidades implicadas para generar confianza y establecer asociaciones sostenibles. Para facilitar los esfuerzos de los WG, los NIH apoyaron conversaciones con personas de sus comunidades en todo Estados Unidos y solicitaron comentarios sobre un conjunto de estudios de casos temáticos y relacionables (apéndice 1; se puede encontrar material más detallado en el sitio web de la Oficina de Política Científica de los NIH¹³ y en la Sección IV).

Se celebraron seis “conversaciones comunitarias” en persona, de cuatro horas de duración cada una. Fueron relativamente reducidos (~25 miembros de la comunidad), con sesiones en grupos para facilitar el debate. Los facilitadores guiaron a los miembros de la comunidad a través de los estudios de casos, que se presentaron en múltiples formatos, mediante texto, ilustraciones y debates aclaratorios. El grupo de trabajo recibió comentarios de comunidades históricamente marginadas de la investigación biomédica. Por ejemplo, un acto (Bronx, NY) se celebró íntegramente en español. También se compensó a los miembros de la comunidad por su tiempo.

A instancias del WG, se celebraron cuatro actos adicionales para incluir perspectivas no recogidas en las

¹¹ Como se indica en la introducción, la mayor parte de las discusiones sobre las preguntas de investigación relevantes se completaron antes de que ChatGPT se pusiera a disposición del público. El Comité desea insistir en que los NIH deben considerar las implicaciones de este tipo de IA generativa a través de la lente de las conclusiones y recomendaciones de este informe (como se indica en el apéndice 2).

¹² El uso compartido de datos federados se refiere a la práctica de utilizar fuentes de datos distribuidas en diferentes ubicaciones para almacenar y controlar los datos, en contraste con el uso compartido de datos centralizado, en el que los datos de múltiples fuentes se agrupan y controlan en una única fuente. Thorogood, A., Rehm, H. L., Goodhand, P., Page, A.J.H., Joly, Y., Boudis, M., Rambla, J., Navarro, A., Nyronen, T.H., Linden, M., Dove, E.S., Fiume, M., Brudno, M., Cline, M.S. y Birney, E. (2021). International federation of genomic medicine databases using GA4GH standards. *Cell Genomics*, 1, 2. <https://doi.org/10.1016/j.xgen.2021.100032>.

¹³ National Institutes of Health, Office of Science Policy (febrero de 2023). *Using Public Engagement to Inform the Use of Data in Biomedical Research, Workshop at University of California, San Diego, 27 de febrero de 2023*. https://osp.od.nih.gov/wp-content/uploads/2023/02/Tab3_NIH_UCSD_Workshop_Community_Conversation_Summary_230227_508.pdf.

conversaciones comunitarias señaladas anteriormente. En tres seminarios web participaron defensores de organizaciones de salud rurales, redes de enfermedades raras y grupos indígenas americanos y nativos de Alaska, y siguieron un orden del día análogo al de anteriores conversaciones comunitarias (apéndice 1). En este caso, los participantes eran defensores profesionales de sus comunidades y, por tanto, representaban una perspectiva más amplia que las personas de las otras conversaciones comunitarias (aunque algunos defensores se identificaron como miembros de las comunidades a las que representan). En San Diego (California) se celebró otro taller organizado por los NIH en el que participaron investigadores, creadores de tecnología y expertos en bioética. El objetivo de este acto era recabar retroalimentación sobre el modo en que los resultados de las conversaciones con la comunidad podrían informar el desarrollo de los programas y políticas de los NIH. Los resúmenes de todos los acontecimientos figuran en la Sección IV.

III. Preguntas de investigación sobre ciencia de datos

A. Introducción y antecedentes

La ciencia de datos sigue acelerando el potencial de descubrimientos y avances en la investigación biomédica. Sin embargo, estas interesantes oportunidades se ven atenuadas por el riesgo que corren las personas, las familias, las comunidades y el público en general, lo que plantea cuestiones fundamentales sobre cómo equilibrar los posibles riesgos y beneficios. Entre los tipos específicos de preguntas y temas de investigación que examinó el Comité figuran los siguientes:

- **Los crecientes volúmenes de datos de salud generados por las personas (PGHD)¹⁴ impulsarán los avances de la ciencia médica, pero siguen existiendo claros obstáculos potenciales:** Cada vez se utilizan más sensores y dispositivos portátiles para vigilar la salud. Las personas pueden capturar e integrar su información de salud personal procedente de innumerables fuentes en aplicaciones, lo que puede capacitarlas para gestionarlas por sí mismas. Además, a los investigadores biomédicos les gustaría incorporar estos datos a la investigación clínica para generar una imagen más holística de las personas y las comunidades. Sin embargo, garantizar la privacidad de los participantes y la seguridad de los datos sigue siendo un reto, y el costo de estas tecnologías emergentes plantea problemas de equidad en salud.
- **Cauteloso entusiasmo por el uso de algoritmos para interpretar los datos de las redes sociales en aplicaciones reales:** los datos de las redes sociales son de dominio público y ha habido mucho interés en aprovecharlos para la investigación biomédica y la salud pública. Por ejemplo, algunas pruebas indican que el NLP y el ML pueden predecir crisis de salud mental para personas y comunidades.^{15,16} Sin embargo, la gente suele sorprenderse de las conclusiones que pueden extraerse de estos datos y no entiende cómo pueden utilizarse. Para evitar la estigmatización de las comunidades y la erosión de la confianza, es necesario seguir trabajando para definir las expectativas de las personas y cuándo y cómo obtener permiso para utilizar sus datos en la investigación biomédica.
- **Mantener a los humanos en el bucle del análisis de datos de investigación biomédica mediante IA:** los modelos y algoritmos de IA incorporan y amplifican los sesgos inherentes a los conjuntos de datos que se utilizan en su desarrollo (es decir, en el “entrenamiento” de los modelos y algoritmos) y a menudo identifican patrones en los datos que no son evidentes para los observadores humanos.^{17,18} Aunque estos algoritmos pueden aumentar la

¹⁴ Como se indica en la nota a pie de página 1, los PGHD son datos recogidos por una persona a través de dispositivos personales diseñados para recopilar métricas relacionadas con la salud (p. ej., tecnologías portátiles y aplicaciones). Puede tratarse de observaciones, opiniones, percepciones, motivaciones, comportamientos, etc.

¹⁵ Birnbaum, M. L., Ernala, S. K., Rizvi, A. F., Arenare, E., R Van Meter, A., De Choudhury, M. y Kane, J. M. (2019). Detecting relapse in youth with psychotic disorders utilizing patient-generated and patient-contributed digital data from Facebook. *NPJ Schizophrenia*, 5, 17. <https://doi.org/10.1038/s41537-019-0085-9>.

¹⁶ Choi, D., Sumner, S. A., Holland, K. M., Draper, J., Murphy, S., Bowen, D. A., Zwald, M., Wang, J., Law, R., Taylor, J., Konjeti, C. y De Choudhury, M. (2020). Development of a machine learning model using multiple, heterogeneous data sources to estimate weekly US suicide fatalities. *JAMA Network Open*, 3, e2030932. <https://doi.org/10.1001/jamanetworkopen.2020.30932>.

¹⁷ Larrazabal, A. J., Nieto, N., Peterson, V., Milone, D. H. y Ferrante, E. (2020). *Gender imbalance in medical imaging datasets produces biased*

interpretación de los datos por parte de los expertos humanos (p. ej., médicos e informáticos), también pueden llegar a conclusiones que tengan consecuencias negativas para determinados grupos. Los investigadores se beneficiarían de unas directrices para examinar, regular e interpretar mejor los resultados de una tecnología específica. Los expertos humanos podrían evaluar la tecnología como un proceso estandarizado e identificar las limitaciones para informar la siguiente fase de investigación e innovación.^{19,20,21}

- **Necesidad de métricas de salud, elementos de datos y modelos normalizados:** un mayor intercambio de datos puede acelerar el progreso de la investigación biomédica al aumentar la potencia estadística y maximizar la utilidad de valiosos conjuntos de datos. Para garantizar la interoperabilidad, los datos y metadatos deben normalizarse en todos los estudios de investigación y jurisdicciones. Existe una variedad de enfoques para estandarizar los datos con fines de comparación, vinculación y agregación.^{22,23} Un reto relacionado es que los investigadores a menudo se basan en métricas de salud tradicionales que pueden no reflejar la complejidad de la fisiología y el bienestar (p. ej., la presión arterial). A medida que se recopilen nuevos tipos de datos personales fuera del sistema de atención médica tradicional, los desarrolladores podrían colaborar con médicos e investigadores para garantizar que los datos sean fáciles de recopilar, se capturen de forma fiable mediante una tecnología asequible y sean predictivos de resultados de salud específicos.

Basándose en las deliberaciones sobre estas consideraciones y los tipos de preguntas de investigación relacionados, el Comité dio prioridad a tres grandes áreas temáticas de investigación para su posterior exploración (tabla 1; figura 1A):

- **Nuevos tipos de datos:** uso de datos novedosos ajenos al sistema de atención médica tradicional.
- **Nuevos tipos de análisis:** Uso de modelos y algoritmos
- **Vinculación y agregación:** vinculación y agregación de conjuntos de datos dispares procedentes de varias fuentes.

Aunque el Comité abordó todo el ciclo de vida de los datos, este marco no abarcaba todas las posibles cuestiones planteadas por la ciencia de datos. También hay solapamiento en los tipos de preguntas de investigación entre áreas temáticas (p. ej., algunos algoritmos novedosos se basan en fuentes de datos novedosas). A continuación se analiza con más detalle cada tipo de pregunta de investigación.

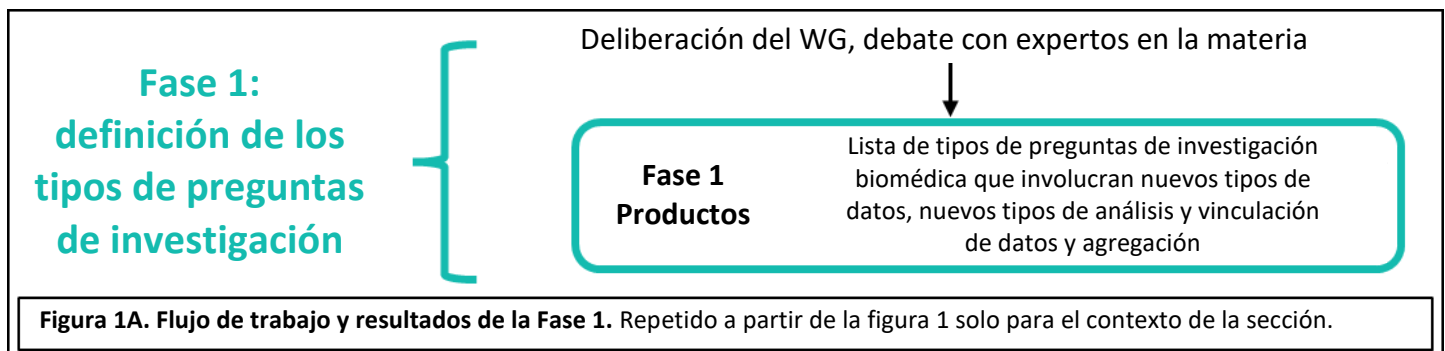


Figura 1A. Flujo de trabajo y resultados de la Fase 1. Repetido a partir de la figura 1 solo para el contexto de la sección.

classifiers for computer-aided diagnosis. PNAS, 117, 12592-12594. <https://doi.org/10.1073/pnas.1919012117>.

¹⁸ Wiens, J., Price, W.N. y Sjoding, M.W. (2020). *Diagnosing bias in data-driven algorithms for healthcare. Nature Medicine, 26, 25-26*

¹⁹ Braatz, J., Rajpurkar, P., Zhang, S., Ng, A. Y. y Shen, J. (2022). *Deep learning-based sparse whole-slide image analysis for the diagnosis of gastric intestinal metaplasia. <https://arxiv.org/abs/2201.01449>.*

²⁰ Holzinger A. (2016). *Interactive machine learning for health informatics: when do we need the human-in-the-loop? Brain informatics, 3, 119-131.*

<https://doi.org/10.1007/s40708-016-0042-6>; Budd, S., Robinson, E. C. y Kainz, B. (2021). *A survey on active learning and human-in-the-loop deep learning for medical image analysis. Medical Image Analysis, 71, 102062. <https://doi.org/10.1016/j.media.2021.102062>.*

²¹ Bodén, A. C. S., Molin, J., Garvin, S., West, R. A., Lundström, C. y Treanor, D. (2021). *The human-in-the-loop: an evaluation of pathologists' interaction with artificial intelligence in clinical practice. Histopathology, 79, 210-218. <https://doi.org/10.1111/his.14356>.*

²² Taber, P., Radloff, C., Del Fiol, G., Staes, C. y Kawamoto, K. (2021). *New standards for clinical decision support: A survey of the state of implementation. Yearbook of Medical Informatics, 30, 159-171. <https://doi.org/10.1055/s-0041-1726502>.*

²³ NeuroLINCS. <http://neurolincs.org/publications/>.

B. Tipos de preguntas de investigación a considerar a través de la participación pública

Uso de datos novedosos ajenos al sistema de atención médica tradicional

El Comité fue más allá de las fuentes tradicionales de datos de salud, como los procedentes de hospitales, historiales médicos de pacientes e información de estudios de investigación primaria. Los ejemplos que figuran a continuación incluyen datos de redes sociales, teléfonos inteligentes, tecnologías portátiles, aplicaciones de salud, datos de consumidores y diversas fuentes sobre SDOH. Los datos procedentes de estas fuentes son únicos porque suelen considerarse datos que *podrían* utilizarse para inferir rasgos de salud, en lugar de indicadores primarios del estado de salud. Sin embargo, en algunos casos, los datos procedentes de dispositivos de consumo pueden ser tan informativos como los obtenidos de dispositivos médicos. Además, los investigadores utilizan cada vez más estos datos en la investigación biomédica y la atención al paciente, difuminando las definiciones tradicionales de datos de salud. El uso de datos que no son de salud para llenar vacíos y, en algunos casos, sustituir a los datos de salud tradicionales es una tendencia especialmente importante a tener en cuenta para la investigación biomédica. Entre los tipos generales y ejemplos específicos de preguntas de investigación en esta área temática se incluyen:

1. ¿Cómo se utilizan los datos personales de salud recogidos fuera del sistema de salud tradicional o del entorno de investigación (p. ej., a través de rastreadores de actividad física, aplicaciones, publicaciones en redes sociales) para estudiar cuestiones relacionadas con la salud y predecir riesgos de salud a nivel individual, familiar, de grupo o de salud pública?

Ejemplos de preguntas de investigación relacionadas con datos de medios sociales (a menudo disponibles públicamente):

- ¿Los datos digitales de Facebook generados por pacientes pueden predecir o detectar recaídas en trastornos psiquiátricos?¹⁵.
- ¿Se pueden utilizar flujos en tiempo real de información secundaria relacionada con el suicidio (p. ej., el tono del lenguaje en las publicaciones y la afiliación a un grupo de usuarios) para estimar con precisión las muertes por suicidio en Estados Unidos en tiempo real?¹⁶.

Ejemplos de preguntas de investigación relacionadas con las tecnologías portátiles (a menudo se presume que los datos son privados, pero el almacenamiento de datos brutos o procesados por parte de terceros puede indicar lo contrario):

- ¿Pueden los micrófonos portátiles emparejados con acelerómetros proporcionar una monitorización cardiopulmonar a largo plazo de los pacientes?²⁴
 - ¿Pueden los sensores inerciales pectorales detectar con precisión una amplia gama de señales fisiológicas (p. ej., parámetros cardíacos y respiratorios) y comportamientos (p. ej., postura, sueño, caídas o deglución) para su seguimiento en tiempo real?²⁵
 - ¿Pueden medirse con exactitud y precisión los niveles de glucosa intersticial para la toma de decisiones clínicas mediante tecnología portátil?²⁶
2. ¿Cómo pueden utilizarse otros datos de consumo y estilo de vida procedentes de fuentes no específicas de la salud (p. ej., sensores en el hogar y datos de tarjetas de crédito y recompensas al consumo) para estudiar cuestiones relacionadas con la salud y predecir riesgos de salud?

Ejemplo de preguntas de investigación:

- ¿Las puntuaciones de crédito pueden definir el riesgo de enfermedad cardiovascular?²⁷

²⁴ Gupta, P., Moghimi, M. J., Jeong, Y., Gupta, D., Inan, O. T. y Ayazi, F. (2020). *Precision wearable accelerometer contact microphones for longitudinal monitoring of mechano-acoustic cardiopulmonary signals*. *NPJ Digital Medicine*, 3, 19. <https://doi.org/10.1038/s41746-020-0225-7>.

²⁵ Rahmani, M. H., Berkvens, R. y Weyn, M. (2021). *Chest-worn inertial sensors: A survey of applications and methods*. *Sensors*, 21, 2875. <https://doi.org/10.3390/s21082875>.

²⁶ Cengiz, E. y Tamborlane, W. V. (2009). *A tale of two compartments: interstitial versus blood glucose monitoring*. *Diabetes technology & Therapeutics*, 11, S11–S16. <https://doi.org/10.1089/dia.2009.0002>.

²⁷ Israel, S., Caspi, A., Belsky, D. W., Harrington, H., Hogan, S., Houts, R., Ramrakha, S., Sanders, S., Poulton, R. y Moffitt, T. E. (2014). *Credit scores*,

- ¿Pueden los asistentes personales de IA (p. ej., Alexa) detectar de forma fiable los problemas de salud a través de los cambios en los patrones del habla?²⁸
 - ¿Pueden aprovecharse los sensores domésticos para crear entornos más seguros para las personas con deterioro cognitivo?²⁹
3. ¿Cómo puede la integración de datos de salud con datos sobre SDOH permitir una mejor estratificación del riesgo de las poblaciones de pacientes y el desarrollo de algoritmos predictivos?³⁰ Los SDOH pueden incluir la situación socioeconómica, el nivel de vivienda, el nivel educativo, los entornos geográficos en los que las personas pasan el tiempo (p. ej., los índices de delincuencia o los contaminantes ambientales en un barrio determinado) y los factores de identidad que favorecen o perjudican el estado de salud.

Ejemplo de preguntas de investigación:

- ¿La evaluación del estado de SDOH en las EHR puede permitir derivaciones adaptadas a las agencias comunitarias disponibles?³¹
- ¿Puede la integración de datos sobre SDOH en las EHR facilitar la predicción del riesgo clínico y la intervención?³²

Uso de modelos y algoritmos

Los recientes avances en análisis, especialmente las técnicas computacionales, están ampliando las fronteras de la investigación biomédica. Por ejemplo, los nuevos métodos de IA, ML y NLP pueden identificar patrones en las EHR de una persona que serían difíciles de descubrir para un clínico. Pero sigue habiendo muchas preguntas e inquietudes sobre cómo se “entrenan” esos modelos y algoritmos y cómo se utilizan en el mundo real. Sin duda, estas consideraciones seguirán siendo importantes y evolucionarán para los investigadores biomédicos y del comportamiento en los próximos 5-10 años. Entre los tipos generales y ejemplos específicos de preguntas de investigación en esta área temática se incluyen:

1. ¿Cuál es el papel de las tecnologías algorítmicas basadas en la informática, como la IA, el NLP (incluidos los grandes modelos lingüísticos como ChatGPT), el ML y el análisis automatizado de imágenes en la investigación biomédica y en la toma de decisiones de salud?

Ejemplo de preguntas de investigación:

- ¿Pueden los algoritmos de aprendizaje profundo (es decir, algoritmos que se actualizan a sí mismos en función de la nueva información) desplegarse para el análisis automatizado eficaz de imágenes médicas con el fin de sustituir la experiencia clínica para el diagnóstico de enfermedades? (p. ej., neumonía³³; cáncer de estómago o intestino¹⁹)
 - ¿Influyen los prejuicios en el desarrollo de algoritmos de forma diferenciada en función de la raza, el sexo o la identidad social o cultural?¹⁷.

cardiovascular disease risk, and human capital. PNAS, 111, 17087-17092. <https://doi.org/10.1073/pnas.1409794111>.

²⁸ Anthes E. (2020). Alexa, ¿tengo COVID-19? *Nature, 586, 22-25. <https://doi.org/10.1038/d41586-020-02732-4>.*

²⁹ Vahia, I. V., Kabelac, Z., Hsu, C. Y., Forester, B. P., Monette, P., May, R., Hobbs, K., Munir, U., Hoti, K. y Katabi, D. (2020). *Radio Signal Sensing and Signal Processing to Monitor Behavioral Symptoms in Dementia: A Case Study. The American Journal of Geriatric Psychiatry, 28, 820-825. <https://doi.org/10.1016/j.jagp.2020.02.012>.*

³⁰ Cantor, M. N. y Thorpe, L. (2018). *Integrating data on social determinants of health into electronic health records. Health Affairs, 37, 585-590. <https://doi.org/10.1377/hlthaff.2017.1252>.*

³¹ Gottlieb, L. M., Tirozzi, K. J., Manchanda, R., Burns, A. R. y Sandel, M. T. (2015). *Moving electronic medical records upstream: incorporating social determinants of health. American Journal of Preventive Medicine, 48, 215-218. <https://doi.org/10.1016/j.amepre.2014.07.009>.*

³² Chen, M., Tan, X. y Padman, R. (2020). *Social determinants of health in electronic health records and their impact on analysis and risk prediction: A systematic review. Journal of the American Medical Informatics Association, 27, 1764-1773. <https://doi.org/10.1093/jamia/ocaa143>.*

³³ Irvin, J. A., Pareek, A., Long, J., Rajpurkar, P., Eng, D. K., Khandwala, N., Haug, P. J., Jephson, A., Conner, K. E., Gordon, B. H., Rodríguez, F., Ng, A. Y., Lungren, M. P. y Dean, N. C. (2022). *CheXED: Comparison of a deep learning model to a clinical decision support system for pneumonia in the emergency department. Journal of Thoracic Imaging, 37, 162-167. <https://doi.org/10.1097/RTI.0000000000000622>.*

- ¿Cuándo y cómo deben introducirse expertos humanos para supervisar y analizar los resultados de los algoritmos y modelos relacionados con la salud a fin de garantizar su fidelidad? (consulte panorama general²⁰; p. ej., experto clínico²¹, supervisión bioinformática³⁴).
 - ¿Los algoritmos pueden determinar el pronóstico de los pacientes al ingresar en el hospital?³⁵
 - ¿Cómo pueden/deben rastrearse y corregirse las estrategias computacionales subyacentes utilizadas por los algoritmos de IA/ML para garantizar la precisión de los diagnósticos clínicos?¹⁸.
 - ¿Qué tipos de sesgo hay en los algoritmos que toman “atajos” para llegar a conclusiones?¹⁹ (Aunque estos algoritmos pueden encontrar soluciones y ofrecer diagnósticos basados en los datos que se les han proporcionado, a menudo no está claro cómo han llegado a esas conclusiones. Los algoritmos pueden estar respondiendo a correlaciones espurias o tomando “atajos” para identificar patrones en los datos)
2. ¿Cómo puede/debe desplegarse el NLP para analizar los datos de los sistemas de salud (p. ej., EHR, datos de seguros médicos, datos de farmacias) con el fin de proporcionar información sobre los síntomas de los pacientes y la clasificación de las enfermedades?

Ejemplo de preguntas de investigación:

- ¿Qué información sobre los síntomas puede encontrarse en los relatos no estructurados de las EHR?³⁶
 - ¿Cómo pueden los grandes modelos lingüísticos exacerbar los sesgos existentes en la investigación médica y la atención clínica?³⁷
 - ¿Cómo influyen las identidades y los antecedentes culturales de médicos y pacientes en la información obtenida mediante NLP a partir de datos no estructurados?

Vinculación y agregación de conjuntos de datos dispares procedentes de múltiples fuentes

A muchos investigadores les gustaría facilitar la reutilización de datos vinculando datos de distintas fuentes. Hasta hace poco, la combinación de diferentes conjuntos de datos introducía importantes problemas de privacidad o disminuía significativamente su utilidad. En el pasado, dos investigadores que querían compartir datos los necesitaban en el mismo formato y debían asegurarse de que al compartirlos no se revelaría accidentalmente información sensible sobre los participantes ni se violaría el consentimiento informado. Con las nuevas herramientas y programas informáticos de protección de la intimidad (p. ej., PPRL), los investigadores pueden vincular conjuntos de datos cumpliendo las normas de protección de la intimidad. Además, se están realizando esfuerzos sistemáticos a gran escala para normalizar los datos de investigación. Aunque persisten algunos problemas de privacidad e interoperabilidad, estas nuevas tecnologías permitirán nuevos descubrimientos en los conjuntos de datos en los próximos 5-10 años. Entre los tipos generales y ejemplos específicos de preguntas de investigación en esta área temática se incluyen:

³⁴ Goodwin, T. R., Demner-Fushman, D. (2020). A customizable deep learning model for nosocomial risk prediction from critical care notes with indirect supervision. *Journal of the American Medical Informatics Association*, 27, 567-576. <https://doi.org/10.1093/jamia/ocaa004>.

³⁵ Kamran, F., Tang, S., Otlis, E., McEvoy, D. S., Saleh, S. N., Gong, J., Li, B. Y., Dutta, S., Liu, X., Medford, R. J., Valley, T. S., West, L. R., Singh, K., Blumberg, S., Donnelly, J. P., Shenoy, E. S., Ayanian, J. Z., Nallamothu, B. K., Sjoding, M. W. y Wiens, J. (2022). Early identification of patients admitted to hospital for covid-19 at risk of clinical deterioration: model development and multisite external validation study. *BMJ*, 376, e068576. <https://doi.org/10.1136/bmj-2021-068576>.

³⁶ Koleck, T. A., Dreisbach, C., Bourne, P. E. y Bakken, S. (2019). Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. *Journal of the American Medical Informatics Association*, 26, 364-379. <https://doi.org/10.1093/jamia/ocy173>.

³⁷ Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P. S. et al. (2021). Ethical and social risks of harm from language models. <https://arxiv.org/abs/2112.04359>.

1. ¿Existen oportunidades para estandarizar los formatos de los datos —o desplegar tecnologías de estandarización— de modo que los datos de diferentes países y sistemas de atención médica puedan agregarse, vincularse y compartirse entre poblaciones?

Ejemplo de preguntas de investigación:

- ¿Puede la investigación observacional a gran escala (facilitada por la estandarización de los formatos de datos entre sistemas de atención médica y países) superar las directrices establecidas y la opinión de los expertos?³⁸
 - ¿En qué medida pueden aplicarse las normas de interoperabilidad actualmente en uso al apoyo a la toma de decisiones clínicas?²²
2. ¿Qué conjuntos de datos dispares (y potencialmente contradictorios) (p. ej., genómica, proteómica, información clínica, imágenes clínicas) pueden vincularse y combinarse mediante agregadores (automatizados) de datos?

Ejemplo de preguntas de investigación:

- ¿Pueden agruparse de forma significativa datos multimodales sobre el cáncer (p. ej., datos de EHR, datos genómicos o datos de imágenes de salud) procedentes de numerosas fuentes para mejorar su utilidad para la comunidad investigadora del cáncer en general?²³.
3. ¿Cómo pueden/deben utilizarse las bibliotecas de salud personales para combinar la información de salud de las personas a través de múltiples flujos de datos diferentes con el fin de informar sobre los resultados de salud?

Ejemplo de preguntas de investigación:

- ¿Cómo pueden combinarse todos los datos conductuales sobre la salud y el estilo de vida de una persona y ser gestionados por ella misma para obtener una imagen precisa de su salud (es decir, un “gemelo digital”)? ¿Pueden utilizarse estos gemelos digitales para mejorar el servicio de atención médica a las personas?³⁹
4. ¿Cómo puede/debe utilizarse la PPRL para combinar datos sobre personas procedentes de múltiples fuentes y con diferentes identificadores para la medicina de precisión y la salud pública?

Ejemplo de preguntas de investigación:

- ¿Puede la PPRL identificar nuevas oportunidades de investigación para anteriores participantes en investigaciones humanas sin sacrificar la privacidad?
 - ¿Puede la PPRL combinar con seguridad datos de salud pública y datos clínicos sin poner en peligro a las personas?
5. ¿Cómo la vinculación y agregación de datos pueden/deben afectar el contexto de la investigación (p. ej., clínico y de salud pública) y el estado de consentimiento de los participantes?

³⁸ Observational Health Data Sciences and Informatics. <https://www.ohdsi.org/>; Schuemie, M. J., Ryan, P. B., Pratt, N., Chen, R., You, S. C., Krumholz, H. M., Madigan, D., Hripcsak, G. y Suchard, M. A. (2020). Principles of Large-scale Evidence Generation and Evaluation across a Network of Databases (LEGEND). *Journal of the American Medical Informatics Association*, 27, 1331-1337. <https://doi.org/10.1093/jamia/ocaa103>.

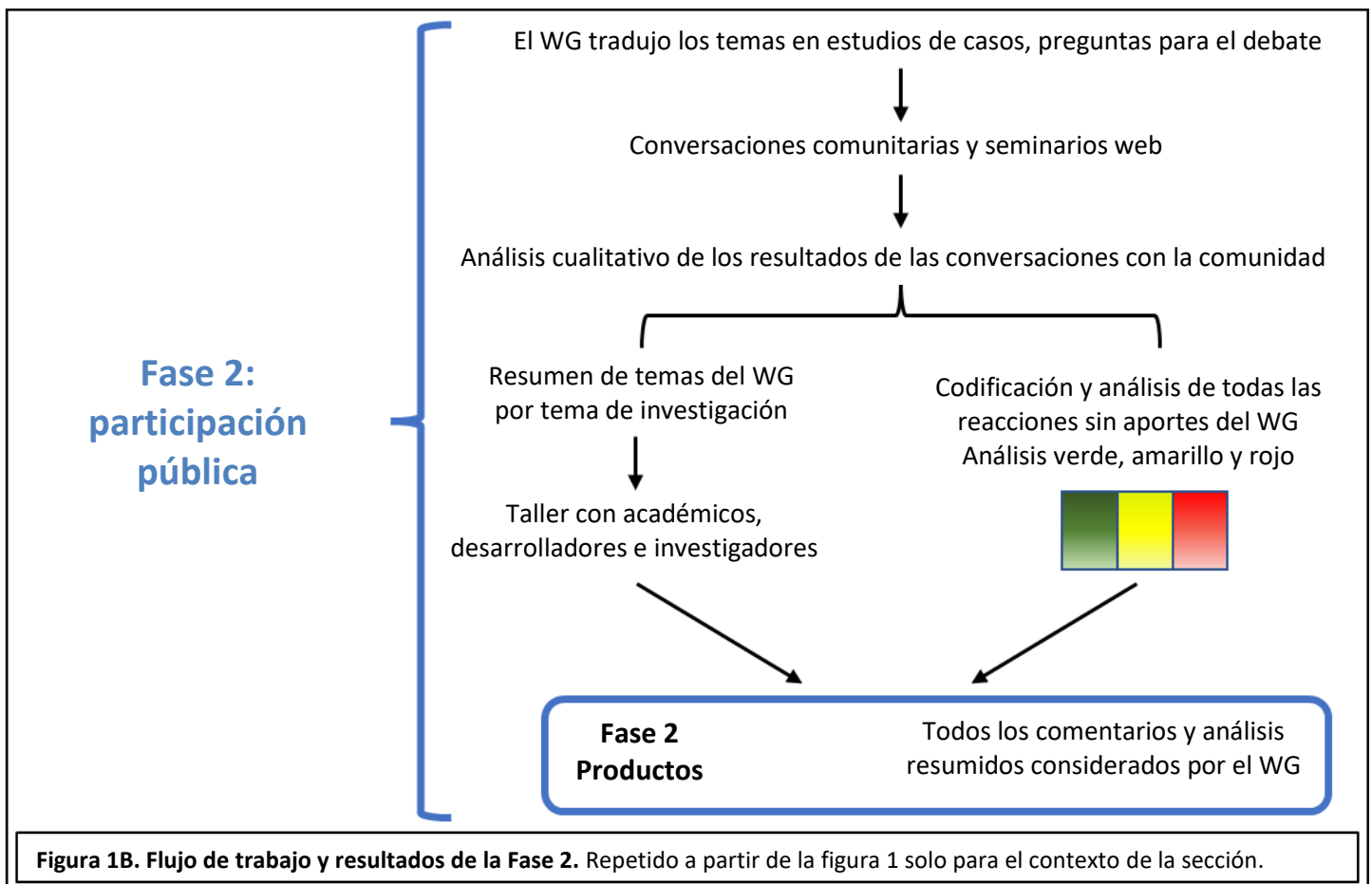
³⁹ Liu, Y., Zhang, L., Yang, Y., Zhou, L., Ren, L., Wang, F., Liu, R., Pang, Z. y Deen, M. J. (2019). A novel cloud-based framework for the elderly healthcare services using digital twin. *IEEE Access*, 7, 49088-49101.

IV. Participación pública para informar sobre las implicaciones de las preguntas de investigación y las tecnologías

A. Recogida de información sobre la participación pública y métodos de análisis

Introducción y antecedentes

La inversión pública en investigación biomédica ha dado lugar a notables avances en la comprensión de la biología fundamental que han transformado la salud humana y la calidad de vida. Como mayor financiador público de investigación biomédica del mundo, los NIH tienen la responsabilidad de garantizar, en la medida de lo posible, que el trabajo que apoyan beneficie a todos e incorpore las preocupaciones y perspectivas públicas. Históricamente, los NIH han recurrido en gran medida a solicitudes de información, reuniones y talleres con investigadores, expertos médicos y estudiosos del ecosistema de la investigación biomédica para recabar la opinión pública. Teniendo en cuenta este encargo, el Comité trató de ampliar su ámbito de actuación y, para ello, los NIH mantuvieron conversaciones con personas de diversas procedencias en comunidades de todo Estados Unidos. Además, los NIH organizaron un taller con académicos de los ámbitos pertinentes (p. ej., expertos en ética y privacidad, científicos de datos, desarrolladores de tecnología en todos los sectores, científicos clínicos y defensores de los pacientes) para reflexionar sobre los comentarios de la comunidad y contextualizar sus perspectivas profesionales dentro de estos debates más amplios (figura 1B; Subsección IV.C).



Los NIH se asociaron con Pyxis Partners, una empresa de defensa de atención médica con facilitadores expertos y relaciones establecidas en comunidades de todo Estados Unidos, para reclutar a miembros de la comunidad y desarrollar materiales para el debate. Se mantuvieron seis conversaciones presenciales de cuatro horas con unos 25 miembros de la

comunidad, cada una de ellas en distintos puntos geográficos. Además, en tres seminarios web se recogieron opiniones de defensores de la salud rural, poblaciones indígenas americanas y nativas de Alaska y defensores de las enfermedades raras (tabla 6). Estas conversaciones se centraron en estudios de casos y preguntas de debate en torno a los tipos de preguntas de investigación definidos en la Fase 1, todo ello desarrollado para maximizar la accesibilidad a los miembros de la comunidad. Para cubrir los distintos elementos de las áreas temáticas de la investigación, se elaboraron dos estudios de caso diferentes para las Áreas Temáticas N.º 1 y N.º 2 (tabla 6).

Las siguientes secciones incluyen un resumen de los debates facilitados, analizados de dos maneras. En primer lugar, el WG identificó los temas clave por área temática de investigación, basándose en las notas y en su experiencia al asistir a estas conversaciones. En segundo lugar, Kendra Nervik, de la Universidad de Wisconsin, Madison, codificó en NVivo los resúmenes de los escritores científicos e identificó los factores que influyen en el apoyo de los miembros de la comunidad a los tipos de investigación. La Sra. Nervik no asistió a las conversaciones comunitarias y, por tanto, ofreció una perspectiva diferente a la de los miembros del WG (tabla 5; tablas 7,8). Los factores identificados por el análisis de la Sra. Nervik están codificados por colores del siguiente modo:

- El verde denota factores que aumentaron el apoyo de los miembros de la comunidad al tipo de investigación. En general, el análisis indica que maximizar estos conceptos beneficiará la fiabilidad y la percepción pública de la investigación.
- El color amarillo denota factores que hicieron que los miembros de la comunidad se detuvieran y buscaran más información antes de decidir si apoyaban el tipo de investigación. Se trata de oportunidades para establecer relaciones de confianza con los posibles participantes en la investigación, anticipándose a sus preguntas e inquietudes y creando recursos para un compromiso continuo.
- El color rojo denota los factores que preocupaban mucho a los miembros de la comunidad y les restaban apoyo al tipo de investigación. En general, la limitación de estos factores estaría más en consonancia con los intereses y valores expresados por los miembros de la comunidad. También brindan a los financiadores e investigadores la oportunidad de comunicar cómo se han tenido en cuenta y abordado o mitigado estas preocupaciones.

Dentro de las categorías codificadas por colores, una mayor saturación de color se correlaciona con más observaciones de un tema en los comentarios de las conversaciones comunitarias. Estas categorías codificadas por colores se solapan en cierta medida en las distintas áreas temáticas de investigación. También existe una fuerte concordancia entre los temas identificados por el WG y el análisis más granular con códigos de colores de la Sra. Nervik. Por último, se eliminó la información potencialmente identificativa (p. ej., citas de miembros de la comunidad) para proteger la privacidad.

B. Conversaciones comunitarias en Estados Unidos

Introducción y antecedentes

De las conversaciones con la comunidad, los seminarios web, las áreas temáticas de investigación y los estudios de casos surgieron tres temas de alto nivel:

- El apoyo de los miembros de la comunidad a las iniciativas de intercambio de datos dependía de quién recibía los datos, de si los receptores se consideraban dignos de confianza y de si habría un retorno significativo de valor para las personas y las comunidades.
- Varios miembros de la comunidad expresaron su deseo de que se introdujeran cambios en el proceso de consentimiento informado para que la gente supiera más sobre cómo se utilizan sus datos (tanto si se recogen primero para la investigación como para otros fines) y algunos querían un mayor control sobre el uso de los datos de ellos o sobre ellos.
- Aunque los miembros de la comunidad expresaron cierto entusiasmo por los nuevos usos de los datos en las tres áreas temáticas, en general querían una mayor transparencia sobre el uso de los datos de investigación, garantías de salvaguardias y razones adicionales para creer que los beneficios merecerían la pena frente a los riesgos potenciales.

Estos temas generales se reflejan a continuación en una descripción más detallada de los comentarios de las conversaciones de la comunidad sobre las áreas temáticas de investigación y los tipos de preguntas de investigación.

Área temática de investigación N.º 1: Investigación y tecnología biomédicas que implican nuevos tipos de datos utilizados fuera del sistema de atención médica tradicional (p. ej., tecnología portátil o redes sociales), como se tipifica en las siguientes preguntas de investigación:

- ¿Cómo se utilizan los datos personales de salud recogidos fuera del sistema de salud tradicional o del entorno de investigación (p. ej., a través de rastreadores de actividad física, aplicaciones, publicaciones en redes sociales) para estudiar cuestiones relacionadas con la salud y predecir riesgos de salud a nivel individual, familiar, de grupo o de salud pública?
- ¿Cómo pueden/deben utilizarse otros datos sobre consumidores y estilos de vida procedentes de fuentes no específicamente de salud (p. ej., sensores en el hogar y datos sobre tarjetas de crédito y recompensas a consumidores) para estudiar cuestiones relacionadas con la salud y predecir riesgos de salud?
- ¿Cómo pueden/deben integrarse los datos de salud con los datos sobre SDOH para permitir una mejor predicción del riesgo y el desarrollo de algoritmos predictivos?

Resumen de los comentarios de la conversación comunitaria sobre el tema de investigación N.º 1

El apoyo de los miembros de la comunidad a compartir y utilizar estos nuevos datos en la investigación biomédica dependía de quién recibía los datos y de si consideraban que esa persona o entidad era digna de confianza. Muchos señalaron que querían recibir cualquier resultado relacionado con la salud de un clínico de confianza con el que tuvieran una relación duradera. Más allá de eso, las respuestas sugieren que existe cierta confianza en el sistema de atención médica, cierta confianza en la investigación y las autoridades gubernamentales (que varía según las personas) y, en general, una confianza muy limitada en la industria. Aunque los miembros de la comunidad expresaron cierto entusiasmo por los nuevos usos de los datos en las tres áreas temáticas de investigación, en general querían una mayor transparencia en lo que implicarían estos usos y garantías de que los beneficios compensarían los riesgos potenciales. Se insistió mucho en la importancia del respeto a la hora de acercarse a las comunidades para generar confianza mediante la transparencia en el uso de los datos de salud personales. Los principales temas de conversación en torno a esta área temática de investigación fueron los siguientes:

- **Confianza:** los miembros de la comunidad conceden gran importancia a la confianza en las personas u organizaciones que reciben y utilizan los datos. A veces se consideraba que las empresas o los investigadores que ganaban dinero compartiendo datos no eran de fiar.
- **Transparencia:** en general, los miembros de la comunidad querían una mayor transparencia sobre el destino de los datos, el uso que se hace de ellos y los objetivos de la investigación que los utiliza.
- **Privacidad:** los miembros de la comunidad querían más garantías de que se protegería su intimidad, especialmente en el caso de los datos utilizados para fines secundarios que podrían estigmatizar o perjudicar de otro modo a las personas o sus comunidades. Por ejemplo, los miembros de la comunidad identificaron la posibilidad de que los datos de las tecnologías portátiles o el uso de sensores domésticos se compartieran de forma que pudieran afectar a los seguros, la salud reproductiva, la situación migratoria o, en general, pudieran estigmatizar a determinados grupos.
- **Control de los usos posteriores:** en general, los miembros de la comunidad querían más control sobre los usos futuros de sus datos, especialmente cuando estos fueran identificables o estuvieran desprotegidos. Por ejemplo, si un estudio utilizaba sensores en el hogar, existía la percepción de que los datos de quienes no participaban en el estudio podían recopilarse y utilizarse para otros fines, con pocas opciones para limitar los usos.
- **Devolución de valor a las personas y las comunidades:** los miembros de la comunidad querían más beneficios en términos de salud para los miembros de la comunidad y las comunidades. Por ejemplo, hubo un gran interés en compartir la información de la investigación (incluidos los resultados) con los proveedores de atención médica para las personas. Algunos miembros de la comunidad también se mostraron escépticos ante la posibilidad de que los datos de las redes sociales fueran lo suficientemente precisos como para traducirse en beneficios para la salud de las personas.
- **Retos del consentimiento informado:** hubo una ligera preocupación por el hecho de que los investigadores obtuvieran datos de salud personales de fuentes públicas o comerciales (p. ej., datos de redes sociales) sin pedir el consentimiento de las personas para hacerlo, así como por la práctica de obtener datos cuando las personas perciben que tienen muy pocas opciones a la hora de consentir que se compartan sus datos (p. ej., tarjetas de recompensa y clickwrap⁴⁰ en servicios o aplicaciones necesarios). Algunos miembros de la comunidad parecían ser conscientes de que estos datos suelen ser propiedad de los desarrolladores o eran de dominio público y preferían dar su consentimiento o, al menos, recibir una notificación del uso de estos datos en la investigación.

⁴⁰ *Clickwrap* es un aviso digital que facilita los procesos de consentimiento ofreciendo a los usuarios la oportunidad de aceptar o rechazar rápidamente las políticas de medios digitales. Observational Health Data Sciences and Informatics. Obar, J. A. y Oeldorf-Hirsch, A. (2018). *The Clickwrap: A political mecanismo económico para fabricar el consentimiento en las redes sociales. Social Media + Society*, 4. <https://doi.org/10.1177/2056305118784770>.

Análisis de los comentarios de la conversación comunitaria sobre los estudios de casos 1 y 2

Para este tema, el análisis cualitativo reveló varios temas que estuvieron presentes en todas las conversaciones comunitarias (tabla 5). Al analizar ambos estudios de caso, los miembros de la comunidad se mostraron más propensos a apoyar la investigación (factores verdes) si sentían que les beneficiaría a ellos y a su comunidad, si los límites de la investigación estaban claros, si había salvaguardias adecuadas para su privacidad, si había oportunidades de controlar el uso de los datos y si esta información era transmitida por alguien que consideraban digno de confianza (especialmente durante el proceso de consentimiento informado). Por otro lado, si los riesgos (p. ej., con respecto a la privacidad de los participantes o la seguridad de los datos) se percibieran como poco claros o superaran a los beneficios, buscarían más información antes de decidir si apoyan la investigación (factores amarillos). Muchos también señalaron que estarían más preocupados cuando se utilizaran datos de personas pertenecientes a clases protegidas (p. ej., menores), personas más fáciles de identificar (p. ej., comunidades tribales en zonas rurales) y aquellas que se identificaran con grupos históricamente perjudicados o marginados en la investigación biomédica. Además, aunque los miembros de la comunidad reconocieron la utilidad de los nuevos tipos de datos en la investigación biomédica cuando están claramente relacionados con las métricas de salud (p. ej., contadores de pasos, monitores de glucosa, dispositivos digitales de salud), muchos consideraron que la relación de los datos de los consumidores con la salud era tenue, que era poco probable que se pudiera extraer algo relevante de los datos de las redes sociales y que había motivos para preocuparse por la pérdida de seguridad y el estigma del uso de esos datos. Por último, aunque algunos miembros de la comunidad reconocieron la necesidad de la colaboración con la industria privada para impulsar la innovación en la investigación, muchos se opusieron a que las empresas ganaran dinero con sus datos (factores rojos). A los miembros de la comunidad les preocupaba el alcance de los datos recopilados y sus posibles usos futuros, así como la posibilidad de que sus datos se vendieran a terceras entidades (p. ej., compañías de seguros), lo que tendría repercusiones negativas para ellos, su familia o su comunidad. Para obtener más información sobre los comentarios por estudio de caso, consulte el apéndice 1.

Tabla 5: análisis cualitativo del apoyo de los miembros de la comunidad al tema de investigación N.º 1

Estudio de caso	Verde: mayor apoyo	Amarillo: incertidumbre o apoyo mixto	Rojo: apoyo disminuido
Estudio de caso 1: Tecnología portátil, Apps, redes sociales y salud cardíaca	Relevancia de los datos de las tecnologías portátiles en la salud.	Transparencia de los procedimientos del estudio	Monetización de los datos
	Familiaridad con las tecnologías portátiles	Privacidad y anonimato	Compartir datos de redes sociales
	Procedimientos para garantizar la privacidad y el anonimato	Compartir información que no es de salud o personal.	Riesgos derivados del uso en fases posteriores
	Explicaciones claras en el consentimiento informado	Compartir datos de las redes sociales	
	Relación de confianza con los investigadores	Agregadores de datos	
	Comunicación clara de los beneficios de la participación	Incluir a las comunidades tribales en la investigación tipos de datos	
		Incluir a menores en la investigación con nuevos tipos de datos	
Estudio de caso 2: sensores a domicilio para personas mayores	Procedimientos para garantizar la seguridad y la confidencialidad	Privacidad y seguridad	Exactitud, eficacia y fiabilidad de los datos
	Transparencia en la recogida y uso de datos	Compartir datos con fines comerciales o de investigación	Alcance de los datos recogidos
	Control de los participantes sobre la recogida de datos	Relación riesgo/beneficio	Recopilación de datos sobre los visitantes
	Compensación a los participantes	Beneficios para las personas y las comunidades	Monetización de los datos
	Consentimiento informado claro		
	Beneficios para la salud de las personas mayores		
	Garantía de prestaciones		

Área temática de investigación N.º 2: Investigación y tecnología biomédicas que implican nuevos tipos de análisis (p. ej., IA, ML y NLP), como se tipifica en las siguientes preguntas de investigación:

- ¿Cuál es el papel de las tecnologías algorítmicas basadas en la informática, como la IA, el NLP (incluidos los grandes modelos lingüísticos como ChatGPT), el ML y el análisis automatizado de imágenes en la investigación biomédica y en la toma de decisiones de salud?
- ¿Cómo puede/debe desplegarse el NLP para analizar los datos de los sistemas de salud (p. ej., EHR, datos de seguros médicos, datos de farmacias) con el fin de proporcionar información sobre los síntomas de los pacientes y la clasificación de las enfermedades?

Resumen de los comentarios de la conversación comunitaria sobre el tema de investigación N.º 2

El apoyo de los miembros de la comunidad al uso de algoritmos y modelos para analizar datos de salud personales dependía de la fiabilidad de los algoritmos a lo largo del tiempo (es decir, si predecían con exactitud lo que decían predecir y podían utilizarse para tomar decisiones de atención médica). Muchos señalaron que si los algoritmos podían demostrar su fiabilidad, les entusiasma la idea de compartir sus datos para recibir diagnósticos más precisos, una mejor atención y abordar problemas de salud pública. Aunque los miembros de la comunidad apoyaban en gran medida el uso de la IA en contextos clínicos, muchos confiaban más en las decisiones de los médicos humanos, basándose en relaciones duraderas y en la confianza existente (haciéndose eco de su preferencia por escuchar a estos asesores de confianza sobre las implicaciones de la recopilación y el uso de nuevos tipos de datos personales para la investigación biomédica en el Área Temática de Investigación N.º 1 anterior). Muchos miembros de la comunidad eran conscientes de la posibilidad de sesgo y mal uso de los algoritmos, en particular de que sus comunidades debían estar representadas en los datos utilizados para desarrollar el algoritmo, y de que el sesgo de los desarrolladores podría afectar a estos modelos y conducir a la estigmatización. En consecuencia, querían saber más sobre la posibilidad de estigmatización y de qué otra forma podrían utilizarse los datos y algoritmos. Los principales temas de conversación en torno a esta área temática de investigación incluyen:

- **Confianza:** aunque estaban abiertos a la idea de confiar en la IA en contextos clínicos, muchos miembros de la comunidad querían escuchar a su propio proveedor de atención médica de confianza para que les proporcionara el contexto y les explicara cómo la IA influía en la toma de decisiones.
- **Transparencia:** a la hora de recibir diagnósticos basados en el análisis de un algoritmo, los miembros de la comunidad dijeron que querían una mayor transparencia sobre cómo se desarrolló el algoritmo, quién lo desarrolló, cómo se llega a las conclusiones y qué poblaciones se estudiaron durante su desarrollo. Muchos miembros de la comunidad querían escuchar tanto a los investigadores como a los proveedores de atención médica para saber más.
- **Precisión y fiabilidad:** los miembros de la comunidad se preguntaron cómo podrían corregirse las conclusiones inexactas de la IA, si esas conclusiones podrían ser perjudiciales y cómo se utilizarían en el futuro esas conclusiones y los datos asociados. Por ejemplo, muchos miembros de la comunidad se mostraron entusiasmados con *el potencial* del uso de algoritmos para analizar los datos de las redes sociales con el fin de ayudar a abordar los problemas de salud mental, pero se mostraron escépticos de que esto se tradujera en intervenciones fiables.
- **Devolución de valor a las personas y las comunidades:** muchos miembros de la comunidad deseaban fervientemente que la investigación se tradujera en beneficios para las personas que participaban en ella y sus comunidades. También querían una mayor participación de las comunidades (incluidos los niños, si procede) a la hora de decidir cómo se utilizarían los datos. Preocupaba que este tipo de investigación pudiera estigmatizar a personas y comunidades.
- **Los riesgos para la intimidad requieren importantes beneficios para la salud:** Muchos miembros de la comunidad percibieron una posible pérdida de privacidad en el uso de la IA, pero expresaron que la compensación merecería la pena si los beneficios para la salud fueran significativos. Por ejemplo, recibir un diagnóstico precoz de cáncer puede merecer el riesgo potencial para la privacidad.
- **Control de los usos posteriores:** en general, los miembros de la comunidad querían tener más control sobre los datos recopilados y analizados por la IA en la investigación, especialmente cuando no veían una conexión de esos datos con el tema principal de un estudio (p. ej., los datos de las redes sociales). Algunos miembros de la comunidad indicaron que estarían encantados de recibir posible información de salud personal que la IA

podiera haber descubierto fuera del ámbito de la investigación para la detección precoz, pero que la posibilidad de que esto pudiera ocurrir debería compartirse de antemano y las personas deberían tener más opciones para controlar cómo se utilizan los datos.

Análisis de los comentarios de la conversación comunitaria sobre los casos 3 y 4

Las conversaciones de la comunidad pusieron de relieve diversas perspectivas sobre el uso de algoritmos y modelos en la investigación biomédica (análisis adicional en el apéndice 1, tabla 7). Muchos de los temas del Área Temática de Investigación N.º 1 que los miembros de la comunidad indicaron que aumentarían su apoyo a la investigación (factores verdes) con nuevos tipos de datos también se mencionaron en relación con el uso de algoritmos y modelos. Los miembros de la comunidad insistieron mucho en la transparencia de la investigación y en las relaciones de confianza establecidas mediante la colaboración con la comunidad. Muchos miembros de la comunidad conocían la IA en forma de algoritmos y modelos, pero carecían de conocimientos técnicos y buscaban una comunicación clara sobre el alcance de su uso en investigación y cómo se desarrollaban. Varios querían comunicarse con asesores de confianza, en particular su proveedor personal de atención médica (en contraposición a los expertos técnicos), para entender cómo ellos y sus comunidades se beneficiarían de la investigación y para tranquilizarse sobre los posibles riesgos para la privacidad, la seguridad y el bienestar de sus comunidades. Estos factores también se reflejaron en algunas de las áreas en las que los miembros de la comunidad expresaron un menor apoyo (factores rojos) al uso de análisis de IA en la investigación. Por ejemplo, si los miembros de la comunidad consideraban que no se podía proteger su privacidad y anonimato, especialmente en la reutilización de los datos más allá del estudio de investigación inicial, era poco probable que apoyaran la investigación. Algunos miembros de la comunidad también eran conscientes de que los algoritmos pueden contener prejuicios y querían asegurarse de que existían mecanismos para evitar que estos análisis los estigmatizaran a ellos o a sus comunidades.

Las actitudes de los miembros de la comunidad variaron en función de cómo se aplicó la IA y del contexto de la investigación (factores amarillos). El estudio del caso 3 se centró en el uso de algoritmos para analizar datos de imagen con el fin de mejorar la detección del cáncer de pulmón. Muchos de los miembros de la comunidad que debatieron este estudio de caso apreciaron y defendieron la utilidad de la IA para “ver” cosas que podrían ser imperceptibles para el ojo del experto humano. Querían más información sobre cómo se protegería su intimidad y se gestionarían y asegurarían sus datos. También tuvieron reacciones encontradas a la devolución de sus resultados individuales, señalando que agradecerían que un clínico de confianza les informara de los resultados de salud personales esperados (es decir, los posibles resultados en el ámbito de la investigación) y que desearían más información cuando los resultados fueran más allá del ámbito del diseño y la descripción iniciales del estudio. En cambio, el estudio del caso 4, que abordaba la recopilación y el análisis de datos de redes sociales públicas para predecir estados de salud mental, suscitó fuertes respuestas a la intersección de los análisis de IA y los datos de redes sociales. A algunos miembros de la comunidad les preocupaba la fiabilidad de lo que la gente comparte a través de las plataformas de redes sociales y si los algoritmos serían capaces de interpretar cualquier matiz (p. ej., cuando las publicaciones intentan exagerar deliberadamente, presentan la información de forma irónica o requieren mucho contexto para entender su significado). Muchos estaban motivados para buscar vías que limitaran el suicidio (en línea con los objetivos de la investigación descritos en el estudio de caso), pero les preocupaba la forma en que los miembros de la comunidad podrían ser señalados a raíz de las conclusiones y cómo se podría estigmatizar a las comunidades.

Área temática de investigación N.º 3: investigación y tecnología biomédicas que implican la vinculación y agregación de datos, como se tipifica en las siguientes preguntas de investigación:

- ¿Existen oportunidades para estandarizar los formatos de los datos —o desplegar tecnologías de estandarización— de modo que los datos de diferentes países y sistemas de atención médica puedan agregarse, vincularse y compartirse entre poblaciones?
- ¿Qué conjuntos de datos dispares (y potencialmente contradictorios) (p. ej., genómica, proteómica, información clínica, imágenes clínicas) pueden vincularse y combinarse mediante agregadores (automatizados) de datos?
- ¿Cómo pueden/deben utilizarse las bibliotecas de salud personales para combinar la información de salud de las personas a través de múltiples flujos de datos diferentes con el fin de informar sobre los resultados de salud?
- ¿Cómo puede/debe utilizarse la PPR para combinar datos sobre personas procedentes de múltiples fuentes y con diferentes identificadores para la medicina de precisión y la salud pública?
- ¿Cómo la vinculación y agregación de datos pueden/deben afectar el contexto de la investigación (p. ej., clínico y de salud pública) y el estado de consentimiento de los participantes?

Resumen de los comentarios de la conversación comunitaria sobre el tema de investigación N.º 3

Los miembros de la comunidad se mostraron escépticos sobre el proceso de vinculación de datos, pero no obstante muchos de ellos apoyaron su uso para avanzar en la investigación. Los miembros de la comunidad eran más propensos a apoyar la vinculación si había una mayor transparencia sobre cómo podrían utilizarse los datos, si recibían información detallada sobre la posibilidad de vinculación durante el proceso de consentimiento informado y si se les podía notificar cuándo y por qué (es decir, con qué fines de investigación) se vincularían sus datos para su uso posterior. Muchos señalaron y apoyaron el uso de la vinculación para acelerar el reclutamiento de miembros de la comunidad por parte de los investigadores para un estudio específico. Algunos miembros de la comunidad, sin embargo, pensaron que esto era incompatible con el espíritu de desidentificación de los datos de los participantes. Al igual que en las áreas temáticas de investigación anteriores, los miembros de la comunidad mostraron su preocupación por la participación de empresas privadas en la vinculación de datos y por el hecho de que muchas entidades diferentes pudieran estar ganando dinero con los datos de los participantes. Algunos miembros de la comunidad consideraron que la vinculación ilustraba cómo el intercambio de datos personales ya está fuera del control de las personas y sus comunidades. El WG tomó nota de los siguientes temas principales de conversación en torno a esta área temática de investigación:

- **Retos del consentimiento informado:** los que apoyaban la vinculación tendían a respaldar la idea de que el consentimiento amplio inicial daba permiso a los investigadores para llevar a cabo la vinculación de datos con datos de salud personales sin identificar, especialmente en el caso de investigaciones que tuvieran un impacto significativo en necesidades de salud especialmente urgentes. Sin embargo, los miembros de la comunidad se mostraron en general escépticos ante los procedimientos de consentimiento de “marcar la casilla” (p. ej., clickwrap) que no describen en detalle los posibles usos futuros, incluida la posibilidad de vinculación. Algunos miembros de la comunidad señalaron su frustración por sentirse como si estuvieran consintiendo en lugar de consintiendo la vinculación en muchos contextos.
- **Transparencia:** aunque algunos de los beneficios de la vinculación eran evidentes para los miembros de la comunidad, en particular para abordar cuestiones relacionadas con la pandemia COVID-19, muchos miembros de la comunidad consideraban que el proceso carecía de transparencia y planteaba demasiados riesgos (p. ej., reidentificación y violación de datos), dada la cantidad de datos disponibles para la vinculación y quién más podría recibir los datos tras la investigación.
- **Privacidad e identificabilidad:** en general, los miembros de la comunidad se mostraron más dispuestos a apoyar la vinculación entre conjuntos de datos completamente anonimizados y agregados. La vinculación de datos más personalizados, como los de las tarjetas de recompensa de las farmacias, se consideró más inapropiada, aunque los miembros de la comunidad reconocieron que podían haber dado su consentimiento cuando se inscribieron (consulte la nota anterior sobre consentimiento frente a aquiescencia).

Análisis de los comentarios de la conversación comunitaria sobre el estudio del caso 5

Como se muestra en la tabla 8 del apéndice 1 (factores verdes), la consideración más importante para el apoyo de los miembros de la comunidad a la vinculación de datos personales de salud sin identificación fue el consentimiento informado. Los miembros de la comunidad apoyaban más la vinculación de datos si habían dado su consentimiento explícito o si se les volvía a contactar para que dieran su consentimiento para una nueva vinculación de datos. Tanto dentro como fuera de los documentos de consentimiento, muchos miembros de la comunidad deseaban más información sobre cuáles de sus datos se vincularían, cómo se compartirían esos datos, cómo utilizarían los investigadores los datos y qué tipos de uso cubriría su consentimiento. En general, solicitaron una comunicación clara y una mayor participación en el proceso de investigación, incluida la notificación de nuevas vinculaciones y la posibilidad de revocar la vinculación de sus datos en fechas posteriores. Los miembros de la comunidad también creían que era importante que los investigadores fueran dignos de confianza y algunos sugirieron que los investigadores establecieran relaciones sostenidas con los miembros de la comunidad y sus comunidades.

Muchos miembros de la comunidad se mostraron algo recelosos ante la vinculación de datos (factores amarillos) debido a la falta de transparencia percibida sobre la posibilidad de una futura vinculación de datos y su uso en la investigación. Los miembros de la comunidad indicaron que se sentirían más cómodos si se ofrecieran detalles en los formularios de consentimiento, en lugar de declaraciones genéricas que simplemente indicaran que se compartirían los datos, y si la información importante se repitiera en notificaciones fáciles de digerir. En relación con esto, varios

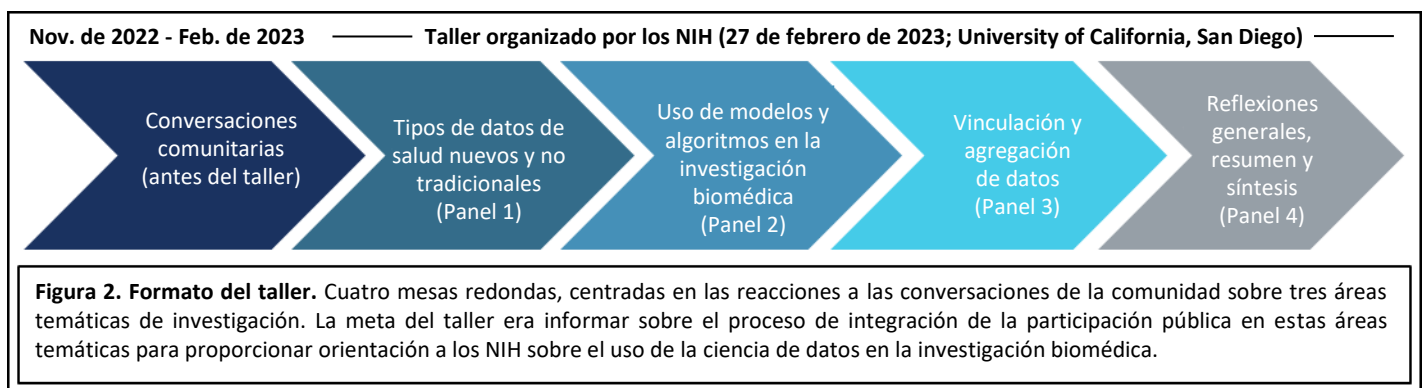
miembros de la comunidad mostraron su preocupación por la redacción de los formularios de consentimiento que permitían amplias categorías de uso futuro de los datos. Aunque muchos señalaron que la investigación futura es difícil de predecir, querían que los formularios de consentimiento caracterizaran la posterior vinculación de datos con la mayor especificidad posible. El proceso de desidentificación también suscitó cierta preocupación. Hasta cierto punto, los miembros de la comunidad se sintieron tranquilizados por el proceso de desidentificación, pero algunos señalaron que la amplia vinculación de datos hacía más probable la reidentificación y sacrificaba las garantías de privacidad. Por último, aunque los miembros de la comunidad apreciaron los beneficios potenciales de la vinculación de datos para los investigadores, querían ver un mayor retorno a los miembros de la comunidad y a sus comunidades (que también fue un tema fuerte en las Áreas Temáticas de Investigación N.º 1 y N.º 2). Sin más retorno de valor, muchos miembros de la comunidad pensaban que los riesgos del uso secundario de los datos enlazados superaban los beneficios potenciales.

De hecho, las mayores preocupaciones de los miembros de la comunidad se centraron en los riesgos percibidos de los futuros usos de los datos enlazados (factores rojos). Los que desaprobaban la vinculación de datos tendían a estar preocupados por la vinculación de datos sensibles y la vinculación de datos que no percibían como directamente relevantes para la salud (p. ej., de tarjetas de recompensa de supermercados). Algunos aludieron al potencial de estigmatización o de juicios negativos, especialmente para las comunidades más fácilmente identificables (p. ej., de tribus o comunidades de enfermedades raras) y se mostraron preocupados por las implicaciones de tales usos. Al igual que en otras áreas temáticas de investigación, los miembros de la comunidad también desaprobaban rotundamente la monetización de sus datos y la posibilidad de que las empresas estuvieran utilizando principalmente la vinculación de datos para ganar dinero.

C. Taller: utilizar la participación pública para informar sobre el uso de datos en la investigación biomédica

Introducción y antecedentes

Los NIH encargaron al NEXTRAC que “convocara consultas con las partes interesadas, incluidos, entre otros, miembros de la comunidad investigadora, grupos de pacientes, expertos en ética y privacidad, científicos de datos, desarrolladores de tecnología de todos los sectores y funcionarios de salud pública.” Las conversaciones con la comunidad descritas en las secciones IV.A y IV.B recogieron opiniones de posibles participantes en la investigación y de grupos de pacientes. Para apoyar las deliberaciones del Comité y ampliar las perspectivas captadas, los NIH organizaron un taller en la Universidad de California, San Diego, para contextualizar los comentarios públicos de las conversaciones comunitarias a través de las lentes de la investigación, la ética, el desarrollo tecnológico y la aplicación a contextos clínicos y de salud pública (página web del evento⁴¹ y grabación de videocast⁴²). El taller constó de cuatro paneles de debate, entre los que se encontraban panelistas de Pyxis Partners que facilitaron las conversaciones de la comunidad para compartir puntos clave y conclusiones, especialistas en ética, investigadores biomédicos, ingenieros tecnológicos y desarrolladores de hardware y software, expertos en salud pública e investigadores clínicos que podrían utilizar estas tecnologías (figura 2).



⁴¹ <https://osp.od.nih.gov/events/nih-workshop-using-public-engagement-to-inform-the-use-of-data-in-biomedical-research/> consultado el 08.05.2023.

⁴² <https://videocast.nih.gov/watch=49084> consultado el 08.05.2023.

Antes del taller, los participantes revisaron los resúmenes de las conversaciones con la comunidad y se les pidió que hicieran una breve reflexión desde su punto de vista personal y profesional. A continuación, los miembros del WG facilitaron las conversaciones sobre los temas de investigación definidos para:

- Debater cómo los diseñadores, desarrolladores y usuarios de datos novedosos consideran el valor y las posibles implicaciones de esta investigación para las personas y las comunidades, y cómo tienen en cuenta estas consideraciones a lo largo del proceso de innovación e investigación
- Contextualizar las conversaciones comunitarias que los NIH celebraron en todo Estados Unidos en un marco bioético más amplio para determinar las lecciones aprendidas
- Integrar los comentarios del público para evaluar las implicaciones para los programas y políticas de los NIH

A continuación se ofrece información resumida de los paneles en referencia a los tres temas de investigación y para motivar comentarios sobre cómo los NIH podrían actuar en el futuro.

Mesa redonda sobre el tema de investigación N.º 1: investigación y tecnología biomédicas que impliquen nuevos tipos de datos utilizados fuera del sistema de atención médica tradicional (p. ej., tecnología portátil y redes sociales).

Los panelistas debatieron sobre la importancia de implicar a las comunidades, ser dignos de confianza, crear capacidad para la investigación comprometida con la comunidad y devolver el valor a los participantes en la investigación que han compartido datos personales a través de las tecnologías emergentes para la investigación biomédica. Los temas debatidos en el Panel 1 incluyen:

- **Recabar la opinión de las comunidades y los participantes en una fase temprana del proceso de investigación:** los participantes animaron a los investigadores y desarrolladores a colaborar con las comunidades desde el principio y con frecuencia, y no solo en lo que respecta a la recopilación de datos o a nuevos tipos de datos.
- **Aprenda cómo definen las comunidades la confianza, la fiabilidad y la transparencia:** la confianza, la fiabilidad y la transparencia fueron temas destacados en las conversaciones de la comunidad y los panelistas animaron a los investigadores a seguir explorando cómo definen estos términos sus poblaciones participantes.
- **Necesidad de una mayor transparencia sobre el uso de los datos, no sólo sobre su origen:** algunos panelistas manifestaron que, con respecto a los “nuevos tipos de datos”, el tipo de datos no importaba tanto como la forma en que pudieran utilizarse.
- **Retorno del beneficio y del valor a los participantes:** muchos panelistas respaldaron la idea de que los participantes y las comunidades deben beneficiarse del intercambio de datos para la investigación biomédica, en particular cuando se consideran las comunidades que han sido históricamente perjudicadas o marginadas por la investigación biomédica.

Mesa redonda sobre el tema de investigación N.º 2: investigación y tecnología biomédicas que impliquen nuevos tipos de análisis (p. ej., IA, ML y NLP)

Los panelistas del Panel 2 también señalaron que los investigadores que llevan a cabo investigaciones biomédicas que implican nuevos tipos de análisis con IA deben comprometerse con las comunidades a lo largo de todo el proceso de investigación y deben ofrecer un mayor retorno al público. Los panelistas debatieron sobre la necesidad de abordar los posibles problemas éticos, jurídicos y sociales con los desarrolladores de tecnología y los investigadores de la industria en una fase temprana de la investigación y la innovación, antes de que surjan problemas. Las conversaciones deben definir los retos y las posibles soluciones para seguir el ritmo de la evolución de los métodos que utilizan la IA en la investigación biomédica y para adaptar mejor la tecnología a la normativa, las políticas y las preferencias del público. Los temas debatidos en el Panel 2 incluyen:

- **Abordar las singulares cuestiones éticas, jurídicas y sociales que plantea la investigación biomédica en la que participa la industria privada:** los panelistas señalaron que muchos miembros de la comunidad no confiaban en las entidades de investigación con ánimo de lucro y recomendaron que los investigadores y desarrolladores de tecnología prestaran más atención a las diferentes normativas con las que trabajan los investigadores de la industria y cómo esas normativas pueden afectar negativamente a los participantes en la investigación (p. ej., al utilizar datos adquiridos a través de acuerdos de licencia de usuario final en lugar del consentimiento informado de la investigación).
- **Necesidad de un compromiso regular de la comunidad en materia de gobernanza de datos para algoritmos:** algunos panelistas recomendaron que el compromiso continuo con las comunidades debería informar el desarrollo de algoritmos para la investigación biomédica con el fin de garantizar la fiabilidad (p. ej., que puedan

utilizarse de forma segura para tomar decisiones de atención médica), la adecuación cultural y para ayudar a abordar el estigma que podría surgir de su uso en la sociedad.

- **Devolución de valor a quienes aportan datos para el desarrollo de algoritmos:** los panelistas sugirieron trabajar más estrechamente con las comunidades para entender qué esperan recibir a cambio de compartir sus datos, ya que probablemente no sea tan sencillo como una compensación económica o la titularidad de la propiedad intelectual.
- **Dar a los ciudadanos más control sobre el intercambio de datos:** algunos panelistas recomendaron que los NIH proporcionen orientación sobre cómo dar a los participantes más control sobre la forma en que se comparten sus datos, así como para facilitar la comprensión de los riesgos, beneficios y compensaciones que implica compartir datos de maneras particulares.
- **Abordar el antiguo reto de mantenerse a la vanguardia del desarrollo de la IA:** algunos panelistas señalaron que las regulaciones, políticas y normas éticas actuales no siempre están bien equipadas para abordar el uso de análisis de datos basados en IA en la investigación biomédica y recomendaron que los NIH y otras agencias lideren el desarrollo de formas de rastrear los algoritmos utilizados en espacios de salud, así como la definición de estándares para evaluar su solidez y la presencia de sesgos.

Mesa redonda sobre el tema de investigación N.º 3: investigación y tecnología biomédicas que implican la vinculación y agregación de datos

La mesa redonda 3 debatió la necesidad de un compromiso respetuoso a la hora de solicitar la participación en investigaciones secundarias, especialmente a través de la vinculación de datos. Esto podría incluir una mejor comunicación durante y más allá del proceso de consentimiento informado y en la prestación de apoyo a las comunidades que soportan mayores riesgos derivados de la vinculación de datos. Los panelistas también reiteraron la necesidad de un compromiso continuo con las comunidades a lo largo de todo el proceso de investigación e innovación. Los temas debatidos en el Panel 3 incluyen:

- **Necesidad de un compromiso temprano y continuo con las comunidades para el uso de datos complejos:** los panelistas sugirieron que las complejidades y sutilezas de la vinculación de datos en la investigación biomédica requieren un compromiso temprano con las comunidades, tal vez incluso en la definición o priorización de las preguntas de investigación, como es coherente con un enfoque de investigación participativa basada en la comunidad.
- **Ilustrar cómo la vinculación y la agregación pueden ser beneficiosas:** algunos panelistas señalaron que la vinculación y agregación de datos puede ser muy beneficiosa para algunas comunidades, pero que esas ventajas deben comunicarse mejor a los posibles participantes, sobre todo a medida que aumenta la complejidad del proceso de vinculación.
- **Incorporar las preferencias tanto individuales como comunitarias:** los panelistas debatieron la tensión ocasional entre las preferencias y objetivos comunitarios e individuales en relación con la participación en la investigación biomédica, y algunos sugirieron que las familias y las comunidades deberían participar para solicitar sus preferencias sobre cómo y cuándo vincular o agregar datos en la investigación biomédica.
- **Proporcionar apoyo y comunicación a las comunidades con mayores riesgos:** algunos panelistas sugirieron que la participación de la comunidad es necesaria para involucrar a las comunidades fácilmente identificables (p. ej., las más pequeñas o rurales) y las comunidades que han sido previamente perjudicadas o estigmatizadas por la investigación biomédica (p. ej., las comunidades de indios americanos y nativos de Alaska) en la toma de decisiones sobre el uso, análisis y difusión de datos.
- **Ser explícito sobre los riesgos para la privacidad que plantea la vinculación de datos:** algunos panelistas sugirieron que los investigadores deberían ser totalmente transparentes sobre los posibles riesgos para la privacidad de la vinculación con los participantes para demostrar su fiabilidad, ya que incluso las tecnologías que se anuncian como “preservadoras de la privacidad” (p. ej., PPRL) pueden aumentar el riesgo de reidentificación.

Reflexiones generales, resumen y síntesis de los panelistas

En el último panel se debatieron las cuestiones generales planteadas en los paneles anteriores. En primer lugar, algunos panelistas sugirieron que los NIH den ejemplo. La industria privada está muy implicada en el desarrollo de nuevos usos y análisis de datos, pero los NIH pueden establecer políticas y normas relativas a la recopilación, gestión, intercambio y vinculación de datos en la investigación biomédica, y eso puede animar a los investigadores del ecosistema más amplio a seguir su ejemplo. En segundo lugar, reconocieron la existencia de una tensión entre el control de los datos y los datos abiertos; la investigación futura y el compromiso de la comunidad deberían transmitir de forma más transparente las compensaciones para las personas y las comunidades entre compartir datos y restringir su uso en la investigación biomédica. A continuación, los panelistas señalaron la necesidad de abordar varios obstáculos estructurales a la

participación en la investigación biomédica utilizando nuevos tipos de datos, análisis y vínculos. Estos problemas van desde la asequibilidad y el acceso limitados a estas tecnologías hasta las limitaciones en la protección de la intimidad. Algunos panelistas instaron a los investigadores a hablar directamente con las personas y comunidades con las que esperan trabajar en cada paso del proceso de investigación, incluso a la hora de dar forma y prioridad a la investigación realizada y tras la finalización de un esfuerzo específico. Una forma de hacerlo sería reforzar el proceso de consentimiento informado a través de múltiples puntos de contacto en la cadena de innovación de la investigación. Un mayor compromiso de la comunidad es también uno de los diversos mecanismos que los investigadores pueden aprovechar para devolver valor a los participantes y sus comunidades. Otras recomendaciones fueron la devolución de los resultados de la investigación y diversas formas de compensación. En resumen, los panelistas sugirieron oportunidades para que los NIH apoyen la investigación en torno a y proporcionen orientación sobre las mejores prácticas para la participación de la comunidad, el proceso de consentimiento informado y la devolución de valor a los participantes. A continuación se resumen los comentarios generales de las conversaciones de la comunidad y del taller como puntos a tener en cuenta para que los investigadores se comprometan con los participantes en la investigación biomédica con datos personales (figura 3). Dado que los usos secundarios de los datos son también una parte esencial de la investigación y la innovación, y que se hizo énfasis en ellos en la participación pública, se señalan de forma diferenciada.



V. Observaciones generales, conclusiones y recomendaciones para los NIH

Introducción y antecedentes

En esta sección se describen las observaciones y conclusiones generales del Comité sobre los temas de investigación en ciencia de datos e investigación biomédica definidos en la Fase 1 y explorados y enriquecidos a través de los comentarios de las conversaciones con la comunidad y el taller realizado durante la Fase 2. El Comité también elaboró recomendaciones para los NIH, extraídas de estas observaciones, conclusiones y otras consideraciones relevantes para las políticas y programas relativos a las cuestiones emergentes en la ciencia de datos.

Dos de las conclusiones del Comité proceden principalmente de las deliberaciones de la Fase 1 y animan a los NIH a seguir apoyando y persiguiendo estas áreas de investigación, cuyos detalles se exponen en la Sección III:

Conclusión 1. Las tecnologías emergentes para capturar, almacenar y analizar datos de salud personales (p. ej., tecnologías portátiles, aplicaciones e inteligencia artificial) ofrecen abundantes recursos para la investigación. Sin embargo, se necesita más trabajo político y ético para aprovechar responsablemente todas sus ventajas.

Conclusión 2. Los nuevos métodos para armonizar y agregar datos procedentes de distintos estudios y de todo el mundo son muy prometedores para la medicina de precisión y la salud pública. Aun así, es necesario introducir mejoras para limitar la variación que puede reducir la significación, la precisión de los resultados y el carácter general de las conclusiones.

El resto de observaciones, conclusiones y recomendaciones proceden principalmente de las deliberaciones de la Fase 2. Al considerar toda la amplitud de temas de las conversaciones comunitarias y el taller público, el Comité identificó tres temas para organizar sus observaciones, conclusiones y recomendaciones:

- **Tema I:** fiabilidad y transparencia en la investigación y respeto por los participantes y las comunidades
- **Tema II:** mejorar el consentimiento informado y el control del intercambio y la vinculación de datos
- **Tema III:** retorno de valor por la participación en investigaciones que impliquen nuevos usos y análisis de datos, incluidos el intercambio y la vinculación de datos

En todos estos temas, hay una variedad de iniciativas -en los NIH y en otros lugares- que informan sobre cómo los investigadores abordan el compromiso público en torno a la ética del uso y análisis de datos personales en la investigación biomédica. Como mayor financiador público del mundo de esta investigación, los NIH están bien situados para identificar los esfuerzos transversales, dilucidar las mejores prácticas y las lecciones aprendidas, y difundir esta información para que los investigadores puedan incorporar eficazmente los comentarios del público en sus prácticas de investigación y gestión de datos. En este sentido, el Comité comparte la siguiente recomendación general:

Recomendación 1: los NIH deberían catalogar, evaluar y, en su caso, coordinar los esfuerzos entre las iniciativas actuales que abordan el compromiso público, las cuestiones éticas y tecnológicas en la ciencia de datos.

El Comité afirma que, a medida que los NIH aborden esta recomendación, así como las que se exponen a continuación por temas, la agencia también debería incorporar métodos para evaluar su aplicación y su impacto en la investigación biomédica. Cuando ha sido útil, el Comité ha añadido un contexto bajo las recomendaciones para orientar las posibles acciones de los NIH.

Tema I: fiabilidad y transparencia en la investigación y respeto por los participantes y las comunidades

Observaciones de las conversaciones con la comunidad: el apoyo de los miembros de la comunidad a los esfuerzos por compartir datos dependía de quién los recibía y de si consideraban que esa persona o entidad era digna de confianza. Muchos señalaron que querían recibir información de salud de un médico de confianza con el que tuvieran una relación duradera. Más allá de eso, los comentarios sugerían cierta confianza en el sistema de atención médica y menos en la industria. Aunque los miembros de la comunidad expresaron cierto entusiasmo por los nuevos usos de los datos en las tres áreas temáticas de investigación, en general querían una mayor transparencia en lo que implicarían estos usos y garantías de que los beneficios compensarían los riesgos potenciales. Se insistió mucho en la importancia del respeto en el acercamiento a las comunidades y en el uso de los datos de salud personales.

Observaciones del taller público: los panelistas animaron a centrarse más en crear y demostrar fiabilidad y mejorar la transparencia no sólo en el uso de los datos, sino en la investigación en general. Sobre lo que sí tienen control los investigadores y las instituciones públicas es sobre si su conducta es digna de confianza. Para generar confianza, muchos panelistas animaron a una mayor transparencia en torno al proceso de investigación, cómo se recogerán, analizarán, gestionarán y utilizarán los datos, y los riesgos y beneficios para los participantes. En todo el ecosistema de la investigación biomédica, y al aplicar la ciencia de datos y las tecnologías emergentes, deben hacerse grandes esfuerzos para desarrollar relaciones duraderas con los participantes y las comunidades a través de un sólido compromiso comunitario. Este compromiso debe adaptarse a comunidades específicas y a los diferentes estadios de la investigación biomédica y el proceso de innovación.⁴³ Esto es fundamental en el contexto de los usos emergentes de los datos personales (individuales y agregados) y sus análisis en la investigación biomédica, ya que presentan un abanico de consideraciones éticas en constante evolución.

Otras consideraciones pertinentes: la opinión pública sobre el uso adecuado de los datos personales en la investigación no siempre se basa en la información más reciente o precisa, puede variar ampliamente y puede evolucionar.⁴⁴ Por ejemplo, las personas pueden no estar familiarizadas con las estructuras reguladoras y los mecanismos de protección de datos en la investigación académica e industrial. La transparencia (o la falta de ella) en torno a las diferentes prácticas en la investigación académica e industrial para el consentimiento, la participación pública y el desarrollo de la IA también puede influir en la percepción de su fiabilidad. En ocasiones, algunos miembros de la comunidad investigadora señalan que puede resultar difícil y requerir mucho tiempo identificar a defensores, enlaces y socios dentro de las comunidades, y que puede ser un reto mantener la fidelidad de la información cuando hay más personas en la cadena de comunicación en torno a un proyecto de investigación.

Conclusión 3. Existe un desajuste entre las expectativas de algunos miembros del público y algunas políticas y programas de los NIH sobre el uso y análisis de datos personales de salud para la investigación biomédica.

Recomendación 2: cuando los NIH invierten en la participación pública, especialmente en torno a las tecnologías emergentes, la agencia debe estar preparada para actualizar los procesos y la infraestructura y para revisar sus programas y políticas de acuerdo con las expectativas públicas y los principios de justicia social, o explicar claramente al público por qué la agencia no puede revisar su curso.

Es importante responder a las opiniones de los ciudadanos. Las expectativas de los ciudadanos pueden variar y, en ocasiones, entrar en conflicto entre las distintas comunidades, y las agencias federales deben, en última instancia, tomar decisiones políticas y de gobernanza basadas en las aportaciones de muchas partes interesadas. Sin embargo, por una cuestión de respeto, para que los miembros del público se sientan escuchados y para generar confianza, los NIH deben estar preparados para traducir las aportaciones del público en cambios significativos. En algunos casos, los NIH pueden trazar un camino diferente al sugerido por las opiniones específicas del público. Cuando esto ocurra, los NIH deberán motivar las decisiones de la agencia (p. ej., como suelen hacer los NIH en respuesta a los comentarios públicos sobre nuevas políticas).

Recomendación 3: los NIH, en consulta con otras agencias federales según proceda, deberían alinear sus políticas sobre datos y gobernanza de la investigación con valores públicos ampliamente extendidos.

⁴³ Unertl, K. M., Schaeffbauer, C. L., Campbell, T. R., Senteio, C., Siek, K. A., Bakken, S. y Veinot, T. C. (2016). Integrating community- based participatory research and informatics approaches to improve the engagement and health of underserved populations. *Journal of the American Medical Informatics Association*, 23, 60–73. <https://doi.org/10.1093/jamia/ocv094>.

⁴⁴ Gerdon, F., Nissenbaum, H., Bach, R.L., Kreuter, F. y Zins, S. (2021). Individual acceptance of using health data for private and public benefit: Changes during the COVID-19 pandemic. *Harvard Data Science Review*. <https://doi.org/10.1162/99608f92.edf2fc97>; Ghafur, S., van Dael, J., Leis, M., Darzi, A. y Sheikh, A. (2020). Public perceptions on data sharing: key insights from the UK and the USA. *Lancet Digital Health*, 2, e444 - e446. [https://doi.org/10.1016/S2589-7500\(20\)30161-8](https://doi.org/10.1016/S2589-7500(20)30161-8); Grande, D., Mitra, N., Iyengar, R., Merchant, R.M., Asch, D.A., Sharma, M. y Cannuscio, C.C. (2022). Consumer willingness to share personal digital information for health-related uses. *JAMA Network Open*, 5, e2144787. <https://doi:10.1001/jamanetworkopen.2021.44787>; Jagsi, R., Griffith, K. A., Jones, R. D., Krenz, C.D., Gornick, M., Spence, R. A., de Vries, R. G., Hawley, S. T., Zon, R. T., Bolte, S., Sadeghi, N., Schilsky, R. L. y Bradbury, A. R. (2019). Effect of public deliberation on patient attitudes regarding consent and data use in a learning health care system for oncology. *Journal of Clinical Oncology*, 37, 3203-3211. <https://doi.org/10.1200/JCO.19.01693>.

No existe un único conjunto de valores públicos establecidos y, especialmente en temas polémicos como el uso adecuado de los datos personales en la investigación biomédica, los valores de las distintas comunidades pueden entrar en conflicto. No obstante, los resultados de las conversaciones de la comunidad y la literatura académica convergen en ciertas cuestiones éticas (p. ej., el interés por la fiabilidad, la transparencia, el respeto y la privacidad).⁴⁵ Cuando surja un mandato claro en las preferencias y expectativas del público, los NIH deberán evaluar y actualizar las políticas y estructuras de la agencia en consecuencia. Para aliviar las cargas administrativas, los NIH deberían intentar armonizar cualquier cambio y comunicación sobre estos temas con las agencias ejecutivas de todo el Gobierno de los Estados Unidos.

Conclusión 4. Es importante consultar a los participantes en la investigación, a las personas cuyos datos se utilizan en la investigación secundaria, a las familias y a las comunidades a lo largo de todo el proceso de investigación biomédica cuando se trata de nuevos tipos, usos y análisis de datos.

Recomendación 4: los NIH deben ayudar a establecer relaciones a largo plazo entre los miembros del público y las comunidades de investigación para facilitar la participación pública sostenida en las cuestiones en evolución relativas al uso, análisis, gestión y reutilización de los datos personales de salud en la investigación biomédica.

El compromiso con las relaciones a largo plazo es fundamental para la confianza y puede sentar las bases de un uso y una reutilización responsables de los datos personales. Las oportunidades y la orientación para una participación significativa en las comunidades en torno a estos temas mantendrán las asociaciones más allá de un estudio individual.

Recomendación 5: los NIH deberían animar a los investigadores a abordar de forma prospectiva en sus programas y propuestas de investigación las consideraciones bioéticas en áreas emergentes de la ciencia de datos que son particularmente importantes para las comunidades participantes, tal y como se refleja en el compromiso público.

Como se indica en las recomendaciones anteriores, los investigadores deben colaborar con las comunidades desde el principio y con frecuencia, incluso a la hora de desarrollar y priorizar las áreas de investigación. De este modo, no solo se obtendrán mejores investigaciones, sino que también se podrán detectar los problemas éticos que plantean la recogida y el uso de datos personales. Los NIH deberían animar a los investigadores a incorporar expertos en bioética y participación de la comunidad en los equipos de investigación y a aprovechar las fuentes existentes de este tipo de conocimientos en sus instituciones de origen o asociadas.⁴⁶

Tema II: consentimiento informado y control del intercambio y la vinculación de datos

Observaciones de las conversaciones con la comunidad: algunos miembros de la comunidad señalaron su frustración por sentir que sus opciones para tomar decisiones sobre sus datos eran limitadas en determinados contextos. Por ejemplo, algunos manifestaron su frustración por el hecho de que los formularios de consentimiento suelen ser largos y difíciles de entender. Además, dado que a menudo están ansiosos por utilizar una tecnología específica o participar en un estudio determinado, su consentimiento para compartir o vincular datos se sintió como una aquiescencia. Muchos miembros de la comunidad deseaban saber más sobre cómo y cuándo podrían utilizarse

⁴⁵ Kim, J., Kim, H., Bell, E. A., Bath, T., Paul, P., Pham, A., Jiang, X., Zheng, K. y Ohno-Machado, L. (2019). Patient perspectives about decisions to share medical data and biospecimens for research. *JAMA Network Open*, 2, e199550. doi:10.1001/jamanetworkopen.2019.9550; Kraft, S. A., Cho, M.K., Gillespie, K. A., Halley, M.C., Varsava, N., Ormond, K. E., Luft, H. S., Wilfond, B. S. y Soo-Jin Lee, S. (2018). Beyond consent: Building trusting relationships with diverse populations in precision medicine research. *American Journal of Bioethics*, 18, 20-23. doi:10.1080/15265161.2018.1431322; Mello, M. M., Lieou, V. y Goodman, S. N. (2018). Clinical trial participants' views of the risks and benefits of data sharing. *New England Journal of Medicine*, 378, 2202-2211. doi:10.1056/NEJMsa1713258; Wilkins, C.H., Mapes, B. M., Jerome, R. N., Villalta-Gil, V., Pulley, J. M. y Harris, P. A. (2019). Understanding what information is valued by research participants, and why. *Health Affairs*, 38, 399-407. doi:10.1377/hlthaff.2018.05046

⁴⁶ Shah, S., Gross, M. y Nebeker, C. (2022). Optimizing ethics engagement in research: Learning from the ethical complexities of studying opioid use in pregnancy. *Journal of Law, Medicine & Ethics*, 50, 339-347. doi:10.1017/jme.2022.61.

sus datos una vez recopilados. Aunque muchos entendían que podían consentir el uso de sus datos más allá de la investigación inicial, muchos también deseaban tener más control sobre esos usos.

Observaciones del taller público: los panelistas señalaron que los enfoques actuales del consentimiento informado para la investigación dificultan que las personas comprendan sus compromisos o sigan comprometidas con un proyecto una vez firmados los acuerdos. Esto es especialmente cierto en el contexto del consentimiento para la investigación más allá del estudio inicial que puede ser complejo o difícil de entender, como cuando se utiliza PPRL. Reconocieron que a los participantes en la investigación les gustaría tener más control sobre cómo se utilizan y reutilizan sus datos y que esto puede estar en contradicción con el movimiento de ciencia abierta en la investigación biomédica. Una de las sugerencias fue que se desarrollara una infraestructura gubernamental para dar a los participantes mayor acceso a un mayor control y seguridad a la hora de compartir sus datos. Una plataforma de este tipo podría utilizarse para mejorar la transparencia en torno al uso de los datos personales, aunque el uso de una infraestructura de este tipo también podría verse dificultado por la posible falta de confianza en la supervisión gubernamental de los datos personales. Varios panelistas sugirieron solicitar la opinión de las comunidades y los participantes como primer paso durante el diseño de la investigación, con énfasis en la comprensión de la pluralidad de enfoques para obtener el consentimiento y cultivar procesos adaptados a las comunidades.

Otras consideraciones pertinentes: hay una serie de limitaciones para mejorar la comprensión y concienciación de los participantes en la investigación a través de los procesos y documentos de consentimiento informado. La investigación sobre la comprensión de los formularios de consentimiento ha demostrado que incluso los formularios concisos y fáciles de leer no suelen conducir a altos niveles de comprensión.⁴⁷ También es extremadamente difícil predecir los usos futuros de los datos personales, lo que limita la capacidad de obtener un consentimiento informado significativo para esos usos. Volver a consentir a los participantes y ofrecerles opciones de control granular o individual de sus datos puede requerir una costosa infraestructura que debe actualizarse periódicamente. Los participantes también pueden sentirse agobiados por las frecuentes peticiones de re consentimiento. Los requisitos legales para el consentimiento también se definen de forma estricta en determinados contextos y se codifican mediante reglamentos supervisados por agencias ajenas a los NIH. Los organismos reguladores también pueden mostrarse reticentes a respaldar un mayor volumen de documentos, notificaciones e infraestructura en torno al proceso de consentimiento, ya que podría resultar complicado armonizar la información y garantizar que todo se ajusta a las normas legales.

Conclusión 5. Se necesita más transparencia en torno al proceso de investigación, cómo se recogerán, analizarán y utilizarán los datos, y los riesgos y beneficios para los participantes en la investigación.

Conclusión 6. La normativa y los procedimientos actuales para el uso de datos de investigación biomédica ofrecen vías para que los datos se reutilicen en la investigación sin consentimiento ni supervisión adecuada.

Conclusión 7. La vinculación de datos en ausencia de un consentimiento informado sólido suscita preocupación por la privacidad de los datos, la transparencia y el respeto a los participantes en la investigación.

Recomendación 6 (que engloba las conclusiones 5-7): los NIH, en consulta con otras agencias federales según proceda, deben desarrollar y difundir las mejores prácticas para informar a las comunidades participantes sobre cómo se gestionan, reutilizan y vinculan sus datos personales de salud con otros datos, especialmente cuando estas mejores prácticas van más allá de los requisitos reglamentarios.

Los NIH han desarrollado recursos para que los investigadores los utilicen en la elaboración de un lenguaje de consentimiento informado para la investigación secundaria con datos y bioespecímenes⁴⁸ y han articulado las

⁴⁷ Beskow, L. M., Lin, L., Dombeck, C., Gao, E. y Weinfurt, K. P. (2016). Improving biobank consent comprehension: a national randomized survey to assess the effect of a simplified form and review/retest intervention. *Genetics in Medicine*, 19, 505 - 512. <https://doi.org/10.1038/gim.2016.157>; Grady, C., Touloumi, G., Walker, A. S., Smolskis, M. C., Sharma, S., Babiker, A. G., Pantazis, N., Tavel, J. A., Florence, E., Sánchez, A., Hudson, F., Papadopoulos, A., Emanuel, E. J., Clewett, M., Munroe, D. y Denning, E. T. (2017). A randomized trial comparing concise and standard consent forms in the START trial. *PLoS ONE*, 12. <https://doi.org/10.1371/journal.pone.0172607>.

⁴⁸ <https://osp.od.nih.gov/wp-content/uploads/Informed-Consent-Resource-for-Secondary-Research-with-Data-and-Biospecimens.pdf> consultado el 31 ago 2023

mejores prácticas para la gestión responsable y el intercambio de datos de participantes indígenas americanos y nativos de Alaska.⁴⁹ Este tipo de recursos son un primer paso importante para orientar y generar confianza. Comunicar información sobre la investigación únicamente a través de un documento de consentimiento informado no basta para generar confianza y, en algunos casos, puede incluso perjudicarla (p. ej., cuando los formularios de consentimiento son demasiado largos o están sobrecargados de jerga). Se necesitan métodos innovadores para mejorar la comunicación tanto dentro como fuera del proceso de consentimiento informado. Las modificaciones para una comunicación más eficaz podrían incluir procesos de consentimiento interactivos, ilustraciones y gráficos que transmitan la información más importante sobre cómo se comparten y protegen los datos personales, un consentimiento estratificado que ofrezca múltiples oportunidades de ver información sobre un estudio y cambiar el propio nivel de participación, y señalización en hospitales y entornos de investigación para indicar cómo podrían utilizarse los datos personales en la investigación y cómo se devuelve el valor a los participantes en la investigación y a las comunidades.⁵⁰

Tema III: devolución del valor de la participación en la investigación y puesta en común y enlace de datos

Observaciones de las conversaciones con la comunidad: el apoyo de los miembros de la comunidad a la investigación dependía de si habría un retorno significativo de valor para ellos, sus comunidades y el público. Algunos miembros de la comunidad señalaron que, históricamente, las comunidades no han recibido información sobre los resultados de su participación en la investigación, ni sobre cómo se han beneficiado ellos u otros, y que esto era especialmente cierto en las comunidades marginadas. Aunque el contexto importaba, muchos expresaron su deseo de saber más sobre los resultados de la investigación en materia de salud personal. Además, algunos señalaron que cuando los investigadores (especialmente en la industria) se beneficiaban de esta investigación, los participantes en la misma debían participar en estos beneficios.

Observaciones del taller público: los panelistas coincidieron en que compartir datos para la investigación debe traducirse en un retorno de valor para los participantes en la investigación y las comunidades. Algunos panelistas señalaron que, con el desarrollo de la IA en la industria, se ha hecho aún más evidente que las empresas se están beneficiando de los datos personales, mientras que quienes aportaron sus datos reciben muy poco a cambio. También señalaron que las distintas comunidades pueden tener concepciones diferentes del tipo de retorno que debe esperarse. Los panelistas debatieron la importancia de proporcionar un retorno específicamente cuando se realizan investigaciones con comunidades marginadas. También sugirieron trabajar más estrechamente con las comunidades para entender qué esperan recibir a cambio de compartir sus datos, ya que no siempre puede tratarse de una compensación económica o de derechos de propiedad intelectual.

Otras consideraciones pertinentes: como demuestran los esfuerzos anteriores, puede ser un reto devolver los resultados de la investigación de forma responsable y de manera que no suponga otros riesgos para los participantes en la investigación.⁵¹ Puede resultar caro y complejo encontrar y compartir de forma segura información de salud privada con los enlaces adecuados para los participantes (p. ej., devolver los resultados a su proveedor de atención médica con suficiente información contextual para transmitir los resultados).

Aunque el retorno de los resultados puede ser una forma de retorno del valor, el valor podría adoptar muchas formas diferentes en función del participante individual o de la comunidad comprometida. Sigue sin estar claro qué “valor” sería más significativo para los participantes y las comunidades (p. ej., financiero, servicios de atención

⁴⁹ Consulte: *NIH Considerations for Researchers Working with AI/AN Communities: Considerations for Researchers Working with AI/AN Communities* | Data Sharing (nih.gov) y <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-22-214.html> consultado el 31 ago 2023.

⁵⁰ Wilbanks, J. (2018). *Design issues in e-consent*. *The Journal of Law, Medicine & Ethics*, 46, 110 - 118. <https://doi.org/10.1177/1073110518766025>; Raj, M., Ryan, K. A., Amara, P. S., Nong, P., Calhoun, K., Trinidad, M. G., Thiel, D. B., Spector-Bagdady, K., de Vries, R. G., Kardia, S. y Platt, J. E. (2023). *Policy preferences regarding health data sharing among patients with cancer: Public deliberations*. *JMIR Cancer*, 9, e39631. <https://doi.org/10.2196/39631>.

⁵¹ National Academies of Sciences, Engineering, and Medicine (2018). *Returning Individual Research Results to Participants: Guidance for a New Research Paradigm*. <https://nap.nationalacademies.org/catalog/25094/returning-individual-research-results-to-participants-guidance-for-a-new>; Wolf, S. M., Ossorio, P. N., Berry, S. A., Greely, H. T., McGuire, A. L., Penny, M. A. y Terry, S. F. (2020). Integrating rules for genomic research, clinical care, public health screening and DTC testing: Creating translational law for translational genomics. *Journal of Law, Medicine & Ethics*, 48, 69–86.

médica o retorno de resultados), en parte porque la base empírica de pruebas es limitada. Además, puede ser difícil asignar valor a las contribuciones de los participantes a un producto o patente de investigación, y determinar una compensación adecuada plantea cuestiones éticas complejas (p. ej., el potencial de coacción y cuestiones de justicia y equidad).

Conclusión 8. Más allá de las promesas potenciales de la medicina de precisión, los participantes en la investigación merecen un mayor retorno de valor para ellos y sus comunidades.

Recomendación 7: los NIH deben aprovechar la investigación previa sobre el retorno de los resultados y apoyar una mayor participación pública para informar sobre las mejores prácticas para el retorno de los resultados de la investigación biomédica utilizando la ciencia de datos emergente y para determinar las necesidades de infraestructura de implementación.

Se han desarrollado varias directrices, marcos e iniciativas para mejorar el retorno de los resultados a las personas y las comunidades.⁵¹ Por diversas razones, estos esfuerzos previos no se han aplicado adecuadamente en los procesos de investigación biomédica. Una mejor difusión, utilización y aplicación de estos recursos ayudaría a los investigadores a satisfacer las preferencias de muchos miembros de la comunidad.

Conclusión 9. Proporcionar a los participantes en la investigación beneficios económicos o un mayor control sobre los datos personales puede reforzar el apoyo de la comunidad a la investigación biomédica realizada por entidades con ánimo de lucro o en asociaciones público-privadas.

Recomendación 8: los NIH deben apoyar la exploración de métodos para proporcionar beneficios y retorno de valor más allá de la devolución de los resultados de la investigación a los participantes en la investigación y sus comunidades, cuando sea apropiado y coherente con la normativa vigente.

Más allá del retorno de los resultados de la investigación, se necesitan más estudios para definir qué forma de retorno podría aportar un valor significativo a los participantes en la investigación y a las comunidades.

El NExTRAC ofrece estas recomendaciones para informar y guiar el desarrollo de futuras políticas y programas en los NIH. El Comité también subraya la importancia de la participación pública a lo largo de todo el proceso de investigación para abordar las necesidades y preferencias emergentes en relación con el uso de los datos de salud personales. La participación pública debe considerarse esencial para una buena investigación y un componente básico del proceso de investigación biomédica, no una tarea accesoria que se realiza cuando conviene.

También es esencial que los NIH, así como los investigadores y las instituciones, piensen en *cómo* se aplicarán estas recomendaciones. Hay que seguir trabajando para definir en qué consiste el éxito de la participación pública, cómo puede medirse y, en general, cómo evaluar el papel de la participación en la investigación biomédica. El NExTRAC fomenta el desarrollo de sistemas y procesos para garantizar que la aplicación de estas recomendaciones conduzca a resultados positivos.

VI. Limitaciones de este esfuerzo y posibles orientaciones futuras

Como se menciona en este informe, el NExTRAC identificó algunas cuestiones que merecen una mayor investigación, pero que no se abordaron en este esfuerzo (p. ej., las aplicaciones de la IA generativa en la investigación biomédica). El Comité subraya aquí otras dos limitaciones y ofrece algunas sugerencias para abordarlas en el futuro. En primer lugar, las industrias farmacéutica y biotecnológica invierten mucho más dinero que los NIH en financiar ensayos clínicos⁵² y las empresas tecnológicas invierten más dinero en el desarrollo de productos que recojan, analicen, almacenen y vinculen datos de salud personales.⁵³ Aunque representantes de las industrias farmacéutica y biotecnológica participaron directamente en los trabajos que condujeron a este informe, este se centra en la investigación financiada por los NIH. Sería necesario un análisis político adicional para evaluar si las conclusiones y recomendaciones de este informe son pertinentes o podrían adaptarse a la investigación clínica financiada por la industria o al desarrollo de tecnologías médicas y de consumo emergentes que puedan utilizarse en dicha investigación. Además, tanto los investigadores biomédicos financiados con fondos públicos como los financiados con fondos privados se beneficiarían de marcos éticos mejor desarrollados para las asociaciones público-privadas. Estos marcos deben abordar cómo se recopilan, comparten, vinculan, analizan y monetizan los datos. El desarrollo de marcos éticos para las asociaciones público-privadas debería ser el centro de futuras investigaciones éticas y políticas.

En segundo lugar, varias cuestiones éticas relacionadas con la recopilación, el uso y la gestión de los datos requieren un mayor análisis académico, la participación de la comunidad y el desarrollo de políticas. En particular, en los debates del NExTRAC no se abordó la justicia reparadora: ¿qué debe ocurrir para restablecer las relaciones entre investigadores y público cuando los investigadores o las agencias utilizan los datos de forma perjudicial, irrespetuosa o explotadora? Mientras que la Recomendación 2 indica que los NIH deben ser sensibles a la falta de alineación entre las políticas y las preferencias del público, no detalla qué debe suceder si o cuando esa falta de alineación conduce a la ruptura de la confianza con las personas y las comunidades. Este informe tampoco aborda los mecanismos de rendición de cuentas para garantizar que los investigadores y las agencias cumplen los compromisos en materia de recopilación y uso de datos. Este informe tampoco describe los mecanismos de procedimiento para determinar el curso de acción cuando las preferencias de la comunidad (o un segmento de ellas) violan los principios de justicia social. En el futuro, la investigación y la participación pública deberán seguir elaborando marcos éticos que incluyan mecanismos de rendición de cuentas sobre la adquisición y el uso de los datos. La justicia reparadora también debe considerarse tanto desde la perspectiva de la bibliografía académica pertinente como de la forma en que los ciudadanos desean participar en los mecanismos de justicia reparadora.

⁵² Zhou, E. W., Jackson M. J., Ledley, F. D. (2023) *Spending on Phased Clinical Development of Approved Drugs by the US National Institutes of Health Compared With Industry*. *JAMA Health Forum*, 4, e231921. doi:10.1001/jamahealthforum.2023.1921;

⁵³ <https://rockhealth.com/insights/h1-2023-digital-health-funding-a-brave-new-lower-funding-world/> consultado el 1 sep 2023;
<https://www.cbinsights.com/research/generative-ai-funding-top-startups-investors/> consultado el 1 sep 2023.

VII. Conclusión

Los recientes avances en ciencia de datos prometen acelerar los descubrimientos en investigación biomédica y mejorar la salud humana. Los nuevos métodos para recopilar datos personales en tiempo real y conectarlos con la experiencia vivida y las exposiciones ambientales podrían hacer posible una medicina de precisión multidimensional y personalizada, y estimular respuestas más rápidas en emergencias de salud pública. Por ejemplo, una persona con hipertensión puede seguir y almacenar simultáneamente en su reloj su frecuencia y ritmo cardíacos, conductancia cutánea, registros neuronales, calidad del sueño y niveles de azúcar en sangre. Sus datos en las redes sociales también pueden ser examinados para identificar otros signos sutiles o predisposiciones a enfermedades cardíacas. Todos estos datos pueden enlazarse y compartirse con los investigadores para obtener nuevos conocimientos sobre la salud humana e identificar a los grupos de riesgo en caso de emergencia.

Estas tecnologías emergentes también plantean cuestiones éticas complejas y cambiantes que pueden alterar el equilibrio entre beneficios y riesgos para las personas, las familias, las comunidades y la sociedad. La sofisticación de los dispositivos y la potencia de cálculo de los algoritmos se están acelerando. Al mismo tiempo, las aplicaciones generalizadas más allá de la investigación y la salud personal han estimulado el debate público sobre la recogida, uso y monetización de los datos personales. Por ejemplo, la idea de recopilar datos personales sobre personas y poblaciones a partir de publicaciones en las redes sociales ha causado alarma en algunas comunidades, concienciando sobre los daños reales y potenciales. Además, algunos han pedido una “pausa” social en el desarrollo y uso de la IA generativa (p. ej., ChatGPT) hasta que se comprendan mejor las implicaciones de estas tecnologías.⁵⁴ El Comité no desea intervenir directamente en esos debates, pero reconoce que parece haber llegado el momento de desarrollar y aplicar nuevas formas de gobernanza. En línea con su marco para evaluar el uso de tecnologías emergentes en la investigación biomédica, el NExTRAC apoya el concepto de que el compromiso público debe informar las nuevas políticas, programas y procesos para gestionar cómo se utilizan y analizan los datos personales en la investigación biomédica.

Los investigadores deben colaborar con las comunidades para plantearse cuestiones como las siguientes: ¿Cuándo deben compartirse y vincularse los datos? ¿Cómo pueden aprovecharse de forma responsable las tecnologías de IA para proporcionar nuevos conocimientos evitando los prejuicios? ¿Cómo pueden los investigadores garantizar la privacidad de los datos al tiempo que obtienen imágenes cada vez más detalladas de las personas y sus familias? ¿Cómo pueden limitar la estigmatización de las comunidades vulnerables y garantizar que esas comunidades se beneficien directamente de la investigación? Algunas de estas preguntas inspiraron el esfuerzo del Comité; la mayoría se ofrecen a los NIH como base para futuras investigaciones a través de las recomendaciones del Comité.

Históricamente, los NIH han recurrido a menudo a métodos de consulta al público que, aunque ampliamente difundidos (p. ej., solicitudes de información, talleres públicos), pueden no ser accesibles para todos. El esfuerzo realizado por el NExTRAC para cumplir con este cometido muestra cómo los NIH pueden mejorar el compromiso público y cultivar relaciones a largo plazo a través de un diálogo más rico y productivo. Al dedicar tiempo a reunirse con miembros del público a través de enlaces de confianza en sus comunidades y adaptar los estudios de casos para maximizar el debate, el Comité recogió innumerables perspectivas sobre cómo se siente la gente acerca del uso y el análisis de sus datos (y los de su comunidad) en la investigación biomédica. Surgieron tres temas: 1) Los NIH deben ayudar a garantizar que el compromiso transparente y sostenido con los miembros del público sustente el desarrollo de políticas y refuerce la confiabilidad de la investigación biomédica; 2) a medida que los investigadores buscan recopilar y reutilizar mayores volúmenes de datos personales, los NIH deben desarrollar y apoyar enfoques para mantener a las personas informadas sobre esos usos; y 3) los NIH deben guiar a los investigadores para encontrar formas de compartir los resultados y devolver el valor a los participantes y sus comunidades.

⁵⁴ Future of Life (22 de marzo de 2023). Pause giant AI experiments: An open letter. <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>.

El Comité reconoce que la profundidad y amplitud de la participación pública deben adaptarse a los distintos tipos de investigación. Cuando esta investigación incorpora nuevos usos y análisis de datos de salud personales, estas prácticas muestran respeto e inclusión de los participantes como socios y pueden reforzar el entusiasmo por los estudios apoyados por los NIH y el ecosistema de investigación biomédica en general. Los NIH deberían utilizar los comentarios de la comunidad, las conclusiones y las recomendaciones de este informe como una hoja de ruta para generar confianza y maximizar la utilidad del dinero de los contribuyentes para su misión de buscar el conocimiento fundamental sobre la naturaleza y el comportamiento de los sistemas vivos y la aplicación de ese conocimiento para mejorar la salud, alargar la vida y reducir la enfermedad y la discapacidad.

VIII. Agradecimientos

En primer lugar, el NExTRAC desea expresar su reconocimiento a las personas de todo Estados Unidos que se unieron a las conversaciones comunitarias. Sus valiosos comentarios han sido fundamentales para este esfuerzo y subyacen a todas las conclusiones y recomendaciones del Comité. En segundo lugar, el Comité desea dar las gracias a los académicos, desarrolladores de tecnología, investigadores clínicos y defensores de los pacientes y la salud que compartieron sus puntos de vista en el desarrollo de la lista de tipos de preguntas de investigación (Sección III) y que reflexionaron sobre los comentarios de la comunidad a través de sus perspectivas profesionales en el taller de la Universidad de California, San Diego (Sección IV). A continuación se enumeran por orden alfabético:

1. Kolbi Brown, Pyxis Partners
2. Mildred Cho, Universidad de Stanford
3. Aaron Coleman, Fitabase
4. Munmun De Choudhury, Instituto de Tecnología de Georgia
5. Luca Foschini, Evidation (anterior), Sage Biosciences
6. Shaun Grannis, Facultad de Medicina de la Universidad de Indiana, Instituto Regenstrief
7. George Hripcsak, Universidad de Columbia, OHDSI
8. Lisa Lehmann, Universidad de Harvard
9. Deven McGraw, Invitae
10. Kendra Nervik, Universidad de Wisconsin, Madison
11. Lucila Ohno-Machado, Facultad de Medicina de Yale
12. Monica Rodriguez, Pyxis Partners
13. Maya Sabatello, Universidad de Columbia
14. Jay Schnitzer, Corporación MITRE
15. Ronnie Tepp, Pyxis Partners
16. Bud Tribble, Apple (anterior)
17. Krystal Tsosie, Universidad Estatal de Arizona
18. Rupa Valdez, Universidad de Virginia
19. Jennifer Wagner, Universidad Estatal de Pensilvania
20. Jenna Wiens, Facultad de Medicina de la Universidad de Michigan

IX. Anexo 1: descripción detallada de la Fase 2

El apéndice 1 incluye más contexto y detalles sobre la Fase 2, incluidos los lugares de las conversaciones comunitarias, los estudios de caso utilizados en cada lugar, el texto de dichos estudios de caso y las preguntas para el debate, así como el análisis rojo-amarillo-verde de los estudios de caso correspondientes a los Temas de investigación N.º 2 y N.º 3. Esto refleja el análisis del Tema N.º 1 de la tabla 5 de la Sección IV.

A. Calendario de conversaciones comunitarias y seminarios web

Se celebraron seis conversaciones en persona en distintos puntos geográficos de Estados Unidos. Además, se celebraron tres seminarios web para recabar opiniones de los defensores de la salud rural, las poblaciones de indígenas americanos y nativos de Alaska y los defensores de las enfermedades raras. La tabla 6 muestra los lugares, las fechas y los estudios de casos analizados en cada sitio.

Ubicación	Fecha	Área temática de investigación N.º 1: tipos de datos novedosos utilizados fuera de lo tradicional Sistema de atención médica	Área temática de investigación N.º 2: nuevas formas de analizar los datos de salud	Área temática de investigación N.º 3: vinculación y agregación de distintos tipos de datos de salud
Harlem, NY (noreste de EE. UU.)	9 nov 2022	Estudio de caso 1: Tecnologías portátiles, aplicación y datos de redes sociales para monitorear la salud cardíaca	Estudio de caso 3: Algoritmo para detectar el cáncer de pulmón (datos de imagen)	Estudio de caso 5: Vinculación de los datos de un estudio sobre demencia con los datos de la COVID
Bronx, NY (Noreste de EE. UU., en español)	10 nov 2022	Estudio de caso 1: Tecnologías portátiles, aplicación y datos de redes sociales para monitorear la salud cardíaca	Estudio de caso 4: Algoritmo en las redes sociales para identificar señales de suicidio en adolescentes	Estudio de caso 5: Vinculación de los datos de un estudio sobre demencia con los datos de la COVID
Alamosa, CO (suroeste de EE. UU.)	5 dic 2022	Estudio de caso 1: Tecnologías portátiles, aplicación y datos de redes sociales para monitorear la salud cardíaca	Estudio de caso 3: Algoritmo para detectar el cáncer de pulmón (datos de imagen)	Estudio de caso 5: Vinculación de los datos de un estudio sobre demencia con los datos de la COVID
Jackson, MS (sureste de EE. UU.)	14 dic 2022	Estudio de caso 2: Sensores domésticos, dispositivos inteligentes para vigilar la salud de un familiar mayor	Estudio de caso 4: Algoritmo en las redes sociales para identificar señales de suicidio en adolescentes	Estudio de caso 5: Vinculación de los datos de un estudio sobre demencia con los datos de la COVID
Dearborn, MI (Medio Oeste, EE. UU.)	18 ene 2023	Estudio de caso 2: Sensores domésticos, dispositivos inteligentes para vigilar la salud de un familiar mayor	Estudio de caso 4: Algoritmo en las redes sociales para identificar señales de suicidio en adolescentes	Estudio de caso 5: Vinculación de los datos de un estudio sobre demencia con los datos de la COVID
Seminario web (defensores de la salud rural, defensores de las enfermedades raras)	24 ene 2023	Estudio de caso 2: Sensores domésticos, dispositivos inteligentes para vigilar la salud de un familiar mayor	Estudio de caso 3: Algoritmo para detectar el cáncer de pulmón (datos de imagen)	Estudio de caso 5: Vinculación de los datos de un estudio sobre demencia con los datos de la COVID
Santa Clara, CA (Oeste, EE. UU.)	31 ene 2023	Estudio de caso 1: Tecnologías portátiles, aplicación y datos de redes sociales para monitorear la salud cardíaca	Estudio de caso 3: Algoritmo para detectar el cáncer de pulmón (datos de imagen)	Estudio de caso 5: Vinculación de los datos de un estudio sobre demencia con los datos de la COVID
Seminarios web (defensores de los indios americanos y los nativos de Alaska)	16 feb 2023 y 17 feb 2023	Estudio de caso 1: Tecnologías portátiles, aplicación y datos de redes sociales para monitorear la salud cardíaca	Estudio de caso 4: Algoritmo en las redes sociales para identificar señales de suicidio en adolescentes	Estudio de caso 5: Vinculación de los datos de un estudio sobre demencia con los datos de la COVID

B. Estudios de casos y análisis de las conversaciones mantenidas en la fase 2

Se elaboraron una serie de estudios de casos y preguntas para debatir con cada comunidad como parte de los compromisos de la Fase 2. A continuación se ofrece el texto completo de los estudios de caso y las preguntas para el debate, aunque se revisaron ligeramente en el transcurso de las conversaciones comunitarias en respuesta a las preguntas y comentarios de los miembros de la comunidad. Debajo de cada estudio de caso hay temas codificados por colores (como se describe en la Sección IV) de las conversaciones sobre esos estudios de caso. Las respuestas a cada estudio de caso también se describen detalladamente en la Sección IV.

Estudio de caso 1: *uso de una tecnología portátil, una aplicación de smartphone y datos de redes sociales para controlar la salud cardíaca*

Le han pedido que participe en un estudio de investigación sobre enfermedades del corazón. Goza de buena salud, pero las enfermedades cardíacas son hereditarias. El estudio de investigación trata de identificar las señales de cuándo o cómo empiezan a ir mal las cosas. Los investigadores esperan que, mediante el seguimiento de varias cosas que se hacen a diario, puedan identificar patrones “saludables” que las personas con riesgo de cardiopatías deberían seguir y actividades “poco saludables” que deberían evitar.

Como parte del estudio, los investigadores le prestan un reloj que puede medir su frecuencia cardíaca y su tensión arterial. También le piden que descargues una aplicación en su teléfono. La aplicación recopila información del reloj, así como otros datos de su teléfono (como su nivel de activ

idad a lo largo del día). También le piden que lleve un registro de los alimentos que come y que introduzca esa información en la aplicación. La aplicación comparte automáticamente toda esta información con los investigadores. Lo único que tiene que hacer es llevar el reloj en la muñeca y el teléfono en el cuerpo o cerca de él (por ejemplo, en el bolsillo o en un bolso que lleve encima), y los investigadores estudiarán los datos recogidos.

Para tener una idea aún más clara de cómo le va, los investigadores también piden permiso para controlar la información que publica en sus cuentas de redes sociales (si es activo en ellas). Los investigadores esperan que, analizando la información procedente de todos estos lugares, podrán comprender mejor los signos precoces de las cardiopatías.

Preguntas para el debate sobre el estudio de caso 1:

- 1. ¿Hay datos personales sobre su salud o de otro tipo (p. ej., redes sociales) que se sentiría más cómodo compartiendo con los investigadores que otros? ¿Qué datos y por qué?*
- 2. ¿Cuál cree que es la mejor manera de que los investigadores utilicen estos datos después del estudio? ¿Quién debe tener permiso para acceder a los datos? ¿Qué se puede hacer con los datos? ¿Quién debe decidir quién puede acceder a los datos?*
- 3. ¿Dejaría que sus hijos participaran en este estudio? ¿Le preocuparía que los niños compartieran alguno de los tipos específicos de datos que los investigadores pretenden recopilar?*
- 4. En este caso, los investigadores han esbozado para los participantes los datos que recogerían, y le han pedido permiso para hacerlo. Sin embargo, algunos de estos datos pueden estar disponibles públicamente o ser recopilados y puestos a la venta por una empresa que fabrica uno de los dispositivos o aplicaciones o por un “agregador de datos” que recopila y empaqueta información de múltiples fuentes. ¿Qué le parecería que los investigadores utilizaran esta información de dominio público o información comprada para crear una imagen sobre su salud sin pedirle permiso?*

Estudio de caso 2: *uso de sensores y dispositivos inteligentes en el hogar para vigilar la salud de un familiar mayor.*

Durante las fiestas del año pasado, sus amigos le hicieron un regalo: unos altavoces “inteligentes”, similares a Google Echo y Amazon Alexa. Los altavoces no solo reproducen música y la radio, sino que también lo escuchan cuando habla en voz alta. Los altavoces están conectados a internet y tienen un asistente virtual con el que puede hablar, hacer preguntas y pedirle que controle otros dispositivos de su casa.

Recientemente, un familiar mayor se ha mudado a su casa. Tiene algunas dificultades para recordar las cosas con claridad y también para mantener el equilibrio, lo que le pone nervioso a la hora de dejarlos solos. Acude a menudo a las

visitas médicas con su familiar y, en una de ellas, el médico les habla a usted y a su familiar sobre un estudio que ayuda a los investigadores del hospital local a aumentar la seguridad de las personas mayores en el hogar.

Los investigadores ofrecen equipar tu casa con una red de sensores que se conectan a los altavoces inteligentes. Por ejemplo, actualmente el Google Nest puede configurarse para subir o bajar automáticamente la temperatura de su casa en función de sus preferencias. En este estudio potencial, algunos de los sensores que traen los investigadores podrían detectar la temperatura y decirle a los altavoces inteligentes que enciendan la calefacción y los ventiladores de su casa cuando llegue a cierta temperatura en el interior. Hay otros sensores que pueden utilizarse para vigilar el estado de salud, como un colchón de seguimiento del sueño, un microchip en el inodoro que detecta marcadores de laboratorio elevados en la orina, una báscula inteligente que mide la pérdida de peso y un espejo digital que nota los cambios de aspecto. Los sensores también pueden analizar patrones de movimiento para detectar si alguien podría lesionarse en caso de caída o despiste. Sin embargo, los investigadores señalan que ninguna de estas funciones es infalible y que la tecnología a veces comete errores.

Los investigadores quieren recopilar datos sobre el habla, el movimiento y las constantes vitales de su familiar mediante una app que les ayudará a identificar posibles causas de problemas de seguridad. Los altavoces que ya están en su casa pueden conectarse a los sensores para recopilar constantemente datos sobre usted y los miembros de su familia y enviarlos a la empresa y a los investigadores para su análisis. Así es como los investigadores controlarían el estado de salud de su familiar. Por ejemplo, los datos podrían diferenciar si su familiar está sufriendo un ictus o simplemente tiene un nivel bajo de azúcar en sangre. La empresa y los investigadores afirman mantener los datos privados y protegidos, pero muchos datos personales se comparten con la empresa a través de la app y los sensores.

Preguntas para el debate sobre el estudio de caso 2 (la tabla 5 de la Sección IV muestra el análisis codificado por colores de los Estudios de Caso N.º 1 y N.º 2):

1. ¿Qué le parece que la empresa obtenga sus datos y los de sus familiares? ¿Cómo se sentiría si otros investigadores pudieran obtener estos datos de la empresa para ayudar a estudiar otras afecciones?
2. Supongamos que vive con otros miembros de la familia y que a menudo invita a familiares y amigos a su casa. Muchos sensores no distinguen entre las distintas personas y se limitan a recopilar datos de todos los que están en la casa en un momento dado. ¿Cómo influiría (afectaría) esto en su decisión de participar en el estudio?
3. ¿Qué preocupaciones tendría (si las hubiera)?
4. ¿Hay factores que aumentarían su probabilidad de participar?
5. ¿Qué importancia tiene para usted comprender los detalles de sus datos y cómo se han analizado? ¿Quiere/necesita conocer todos los detalles? ¿Quién preferiría que le ayudara a analizar (comprender/interpretar) sus datos (p. ej., su médico, un investigador u otra persona)?

Estudio de caso 3: uso de un algoritmo sobre datos de imagen para detectar cáncer de pulmón en estadios tempranos.

El cáncer de pulmón es una enfermedad compleja. Hay personas que fuman a diario durante toda su vida adulta y siguen sin tener cáncer. Otros viven toda su vida en entornos con aire limpio, evitan fumar y no tienen antecedentes familiares de enfermedad pulmonar, pero por alguna razón desarrollan cáncer de pulmón. Aún queda mucho por aprender, y los investigadores siguen buscando factores de riesgo precoces para evitar que las personas contraigan cáncer de pulmón o para diagnosticar ese cáncer en una fase temprana que permita un tratamiento precoz.

En su revisión anual con el médico, ve un folleto en el que se buscan voluntarios sanos de su grupo de edad para participar en un estudio de investigación que busca factores de riesgo precoces de enfermedad pulmonar. Es un compromiso de tiempo mínimo. Usted tiene antecedentes familiares de cáncer de pulmón y, aunque sus familiares afectados eran grandes fumadores y usted nunca ha fumado, decide participar en el estudio.

Cuando habla con el reclutador del estudio, le informa que los investigadores están desarrollando una forma de comparar los historiales médicos de personas sanas con los de personas con cáncer de pulmón, utilizando inteligencia artificial (IA). Como parte del estudio, se le someterá a un procedimiento de detección (tomografía computarizada (TC)). Los investigadores compararán su escáner con los de pacientes sanos que nunca tuvieron cáncer y con los de pacientes

con distintos tipos y estadios de cáncer de pulmón. Los investigadores le informan que tiene la opción de recibir los resultados del estudio, y le concertarán una cita con un neumólogo para que los analice a su conveniencia.

Unos meses después de inscribirse en el estudio, acude al neumólogo, que le da los siguientes resultados:

- Hay algunos factores en su historial médico que se consideran factores de riesgo. Aunque usted no ha fumado, sus padres sí lo hicieron en su casa durante toda su infancia. Además, se descubrió que usted trabaja y ha vivido cerca de algunas fábricas industriales que liberan sustancias químicas que pueden causar cáncer.
- El médico le informa que los escáneres se analizaron con un algoritmo de IA diseñado específicamente para detectar el cáncer de pulmón en estadio inicial. Sin embargo, el médico no puede decirle mucho más sobre cómo funciona el algoritmo. El médico solo conoce el resultado: no se ha detectado ningún tumor. El médico no tiene más preocupaciones en este momento. Sin embargo, dados los factores de riesgo descubiertos, recomiendan revisiones más frecuentes (anuales frente a cada 5 años).

Preguntas de debate sobre el estudio de caso 3:

1. Los algoritmos (es decir, los cálculos informatizados) tienen en cuenta mucha información que ayuda a su médico. ¿Es más o menos probable que confíe en los diagnósticos y orientaciones de su médico que en los de una herramienta basada en IA? ¿Qué más le gustaría saber, si es que hay algo más, sobre el proceso de escaneo de las publicaciones en las redes sociales por parte de los investigadores? ¿Hay factores que aumentarían o disminuirían su confianza en su uso?
2. ¿A quién le gustaría escuchar para que lo ayudara a comprender este tipo de información de salud predictiva o diagnóstica?
3. ¿Qué le gustaría saber sobre los algoritmos utilizados?
4. ¿Hay factores que aumentarían o disminuirían su confianza en su uso?
5. ¿Le preocupa la privacidad?
6. ¿Cómo equilibra las ventajas de conocer esta información de salud con las preocupaciones sobre la privacidad que pueda tener?

Estudio de caso 4: uso de un algoritmo sobre datos de redes sociales para identificar indicios de suicidio en adolescentes.

Supongamos que vive en una ciudad en la que se ha producido recientemente un aumento de las muertes de adolescentes. La pandemia ha sido dura para todos: la gente se ha quedado atrapada en sus casas, la economía ha sufrido un duro golpe y la presión sobre los estudiantes de secundaria y bachillerato con la escuela virtual es mayor que nunca. Como estos adolescentes llevan tiempo sin poder relacionarse en persona, se expresan y conectan sobre todo a través de aplicaciones de redes sociales (p. ej., Instagram, TikTok, Facebook). Sus vecinos tienen todo tipo de teorías sobre lo que podría estar causando tantas muertes entre las personas más jóvenes, pero es difícil saberlo con certeza. En cualquier caso, todo el mundo está de acuerdo en que es un problema y hay que hacer algo.

Los investigadores han desarrollado una forma de escanear e interpretar rápidamente los mensajes de las redes sociales, incluidos vídeos, imágenes, sonido y texto. Los funcionarios de salud pública de su ciudad se han puesto en contacto con estos investigadores para determinar si este recurso podría utilizarse para llegar a las causas subyacentes de las muertes de adolescentes, identificar signos de suicidalidad en los adolescentes y desarrollar formas de resolver este problema. Los investigadores aceptan poner a prueba su proceso en su comunidad, escanearán las publicaciones en las redes sociales y utilizarán los datos para obtener información como la renta media y el nivel educativo del vecindario, así como los índices de delincuencia.

Sus funcionarios públicos han organizado un ayuntamiento con estos investigadores para dialogar con usted y sus vecinos sobre los pros y los contras del uso de esta tecnología.

Preguntas para el debate sobre el estudio de caso 4:

1. ¿Cuál es su primera pregunta para los investigadores? ¿Para los funcionarios de salud pública?
2. ¿Qué más le gustaría saber, si es que hay algo más, sobre el proceso de escaneo de las publicaciones en las redes sociales por parte de los investigadores? ¿Hay factores que aumentarían o disminuirían su confianza en su uso?

3. Si acepta colaborar con los investigadores, estos querrán publicar sus resultados y compartir los datos para que puedan ser reutilizados en futuros estudios o por otros investigadores. Normalmente, antes de publicarlos o compartirlos, se elimina de los datos toda información que permita identificarlos, aunque esto puede limitar la capacidad de utilizar los datos para ayudar a las personas. ¿Basta con eliminar los datos identificativos? ¿Hay otras condiciones que considere necesarias antes de dar su consentimiento para que se publiquen y compartan estos datos sobre sus hijos y los de sus vecinos?
4. ¿Su respuesta sería diferente si estuviéramos hablando de adultos y no de niños?

Tabla 7: análisis cualitativo del apoyo de los miembros de la comunidad al tema de investigación N.º 2			
Estudio de caso	Verde: mayor apoyo	Amarillo: incertidumbre o apoyo mixto	Rojo: apoyo disminuido
Estudio de caso 3: algoritmo para detectar el cáncer de pulmón	Transparencia en la creación del algoritmo	Privacidad y seguridad	Inexactitud de los algoritmos
	Transparencia en el financiamiento de algoritmos	Obtención de resultados	Datos no representativos de la población
	Precisión o validez del algoritmo		Riesgos derivados del uso secundario de datos
	Datos representativos de la población		
	Confianza tanto en los médicos como en los algoritmos		
	Relación de confianza con el médico personal		
Transparencia sobre los usos			
Estudio de caso 4: algoritmo de redes sociales para identificar indicios de suicidio	Comunicación clara sobre el alcance de la investigación y el uso futuro de los datos	Amplio alcance de los datos de las redes sociales	Causar estigma
	Transparencia y honestidad sobre la investigación	Recogida de datos de redes sociales	Riesgos derivados del uso secundario
	Aportaciones comunitarias a la investigación	Seguridad de los menores	Falta de privacidad y anonimato
	Explicaciones claras sobre el funcionamiento de los algoritmos	Garantizar el anonimato de los datos de las redes sociales	
	Garantías de privacidad y seguridad	Relevancia de los datos de las redes sociales para el suicidio	
	Consentimiento informado explícito y claro	Información estigmatizante	
	Agencia participante sobre la recopilación de datos y la participación	Datos inexactos o poco fiables de las redes sociales	
		Relación riesgo/beneficio	
	Beneficios poco claros		

Estudio de caso 5: vinculación de datos de un estudio sobre demencia con COVID y otros datos, junto con el uso de una base de datos de vinculación para facilitar el reclutamiento para estudios relacionados con la demencia.

Supongamos que decide participar en un estudio sobre la demencia en busca de factores de riesgo y tratamientos tempranos. No tiene problemas de salud y se considera una persona bastante sana. Sin embargo, ha visto a amigos y familiares sufrir demencia y espera vivir bien a medida que envejezca, así que acepta participar. Lo único que se le pide es que responda a un cuestionario básico sobre sus hábitos diarios. También acepta compartir sus datos de salud personales para contribuir a este y otros proyectos de investigación. Usted marca una casilla que dice: “Acepto que mis datos se compartan para futuras investigaciones biomédicas”

Supongamos que, al margen de este estudio, usted proporciona sangre y saliva como parte de las pruebas rutinarias de COVID-19. Se enteró de que tenía COVID-19, pero sus síntomas eran leves. Tuvo goteo nasal y dolor de garganta que desaparecieron al cabo de una semana. Vuelve a marcar una casilla que dice: “Acepto que las muestras y los datos asociados de mi prueba COVID-19 se compartan para futuras investigaciones biomédicas”.

Meses después de finalizar el estudio sobre la demencia y las pruebas COVID-19, recibe una llamada de otro equipo de investigación. Le preguntan si desea participar en un estudio de Long COVID. Están intentando determinar cómo la infección previa por COVID-19, que puede interferir en la función cerebral (desde la pérdida del olfato y el gusto hasta la “niebla cerebral”), puede afectar a los primeros indicadores de demencia. Pero, ¿por qué lo llaman?

Los investigadores explican que recibieron información de una base de datos que relaciona datos de distintos estudios. La base de datos indicaba que usted había proporcionado sangre y saliva como parte de las pruebas rutinarias de COVID-19. Aunque eso fue hace tiempo y no para investigación, usted había dado su consentimiento para que esos especímenes se conservaran para su uso en futuras investigaciones. Resulta que, cuando usted participó en el estudio sobre la demencia, la base de datos vinculó automáticamente sus datos de la encuesta sobre la demencia con sus datos de las pruebas COVID-19. Esos datos están ahora disponibles para que otros investigadores los soliciten para utilizarlos en otros estudios. Los investigadores de Long COVID pudieron utilizar la base de datos para ver que usted había participado anteriormente en investigaciones relevantes para su nuevo estudio.

Dado que la demencia podría estar relacionada con Long COVID, los investigadores están muy interesados en su participación. Los investigadores ya han relacionado los datos de la encuesta sobre demencia con sus datos de las pruebas COVID-19. Le hacen saber que aún queda mucho por aprender sobre el impacto a largo plazo de la infección por COVID y que la obtención de más datos personales sobre su salud servirá de base para nuevas intervenciones y terapias. Además, si los investigadores encuentran algo significativo en sus datos personales, tendrá la opción de comentar estos resultados con el médico de su equipo.

Preguntas para el debate sobre el estudio de caso 5:

1. ¿Cómo se sentiría si lo reclutaran de esta manera para el estudio Long COVID?
2. ¿Le parece bien que se vinculen los datos de diferentes estudios como éste? ¿Le parece bien que los investigadores vinculen y utilicen otros datos suyos que usted haya consentido que se recopilen y almacenen (p. ej., de las pruebas COVID-19)?
3. (a) Por ejemplo, ¿qué pasaría si se vinculasen los datos de una compra con tarjeta de recompensa en una farmacia? (b) ¿Qué pasaría si los datos procediesen de Medicare o Medicaid o de alguna otra base de datos de salud que contenga información personal? (c) ¿Qué pasaría si los datos procediesen del censo nacional o de bases de datos sobre delincuencia local, contaminación, vivienda y otros factores que afectan a las comunidades?
4. En este caso, usted consintió en compartir su información para futuras investigaciones biomédicas, pero no específicamente para la vinculación de datos. ¿Esperaría que los investigadores tuvieran que volver a pedirle permiso antes de obtener datos vinculados sobre usted?
5. ¿Qué más cree que habría que hacer, si es que hay que hacer algo, para vincular sus datos de estas formas?

Tabla 8: análisis cualitativo del apoyo de los miembros de la comunidad al tema de investigación N.º 3			
Estudio de caso	Verde: mayor apoyo	Amarillo: incertidumbre o apoyo mixto	Rojo: apoyo disminuido
Estudio de caso 5: vinculación de datos	Consentimiento informado explícito	Transparencia sobre la posibilidad de vinculación	Vinculación con datos sensibles
	Autorización renovada para nuevos usos	Detalles sobre la vinculación	Vinculación con datos que no son de salud.
	La agencia del participante sobre los nuevos usos	Coherencia entre la vinculación y el consentimiento inicial	
	Comunicación clara de los usos futuros de los datos	Anonimato	
	Comunicación clara de la recogida de datos	Seguridad y privacidad	
	Capacidad de retirada	Exactitud y fiabilidad de los datos enlazados	
	Relación de confianza con los investigadores	Traducción a prestaciones	
	Notificación de nuevos usos de datos	Relación riesgo/beneficio	
Transparencia sobre la investigación	Riesgos derivados del uso secundario		

X. Apéndice 2: la IA generativa como caso práctico para las recomendaciones de los WG

A. Introducción

La IA generativa y, en particular, los modelos de NLP estaban en fase de desarrollo y uso al inicio de este esfuerzo en noviembre de 2021. En aquella época, las máquinas de aprendizaje lógico (Logic Learning Machine, LLM) se utilizaban para extraer información de las EHR y los mensajes de las redes sociales, pero estas tecnologías aún no habían irrumpido en la vida pública ni en los medios de comunicación populares. En consecuencia, el Comité tomó nota de estas tecnologías y de sus posibles aplicaciones al abordar la primera parte del encargo (Sección III).

A principios de noviembre de 2022, OpenAI publicó una primera versión de ChatGPT.⁵⁵ Se trata de la primera interfaz disponible para que los usuarios exploren la utilidad de una LLM, ampliando los debates sobre esta tecnología desde las redes locales de la industria y el mundo académico hasta los medios de comunicación populares. Esto generó provocadores debates sobre la posible aplicación de las LLM en todas las facetas de la sociedad, incluido el uso de las LLM en la investigación biomédica. Dado que las LLM como ChatGPT aún no habían aparecido en el espíritu público de entonces, no fueron el foco principal de la Fase 2. Para complementar la discusión de modelos y algoritmos en los Estudios de los Casos N.º 3 y N.º 4, el NExTRAC eligió explorar la aplicabilidad de sus recomendaciones a través del siguiente Caso de Uso como prueba de cómo las recomendaciones de este informe podrían aplicarse a las LLM.

B. Caso práctico

Consideremos una propuesta de investigación teórica presentada como parte de un hipotético programa financiado por los NIH sobre IA generativa y el retorno de los resultados clínicos a los pacientes:

Programa hipotético de los NIH: se solicitan solicitudes de financiamiento sobre el uso de LLM que cumplan la HIPAA para hacer llegar los resultados de los análisis clínicos a los pacientes.

Estudio de investigación hipotético propuesto: Un ensayo controlado aleatorizado para determinar cómo perciben los participantes en la investigación la devolución de sus resultados de una tomografía computarizada (TC) a través de sus portales de pacientes en una de dos formas: un informe estándar basado en las prácticas radiológicas actuales y que utiliza un lenguaje destinado a los médicos clínicos frente a un informe sobre la exploración que ha sido traducido por una LLM que cumple con la HIPAA para que sea más legible y fácil de interpretar por los participantes en la investigación.

La tabla que figura a continuación delinea una forma en que las recomendaciones del NExTRAC (Resumen Ejecutivo y Sección V) podrían interpretarse para esta pregunta de investigación de prueba:

Recomendación	Aplicación a un hipotético 1) Programa NIH y 2) Estudio de investigación
1. Los NIH deberían catalogar, evaluar y, en su caso, coordinar los esfuerzos entre las iniciativas actuales que abordan el compromiso público, las cuestiones éticas y tecnológicas en la ciencia de datos.	1) Los NIH deberían catalogar los esfuerzos existentes que utilizan LLM en la investigación biomédica y el retorno de los resultados de la investigación. En la medida de lo posible, estos esfuerzos deberían ser evaluados en busca de las mejores prácticas y deberían desarrollarse orientaciones generalizables y puntos a tener en cuenta como recursos para la comunidad investigadora. Estos recursos deben estar vinculados al anuncio de financiamiento del Programa y se debe animar a los investigadores a que hagan referencia a ellos en las propuestas. 2) Entre los esfuerzos pertinentes que el investigador podría citar se incluyen:

⁵⁵ OpenAI (30 de noviembre de 2022). Introducing ChatGPT. <https://openai.com/blog/chatgpt>

	<ul style="list-style-type: none"> • Uso generalizado de la IA en medicina • Explicación de los modelos y resultados de la IA • Consentimiento informado en la investigación clínica • Obtención de datos y posibles sesgos de las LLM. • Participación de la comunidad en la investigación clínica
	<ul style="list-style-type: none"> • Proyectos éticos para informar y explicar los resultados de las imágenes
<p>2. Cuando los NIH invierten en la participación pública, especialmente en torno a las tecnologías emergentes, la agencia debe estar preparada para actualizar los procesos y la infraestructura y para revisar sus programas y políticas de acuerdo con las expectativas públicas y los principios de justicia social, o explicar claramente al público por qué la agencia no puede revisar su curso.</p>	<ol style="list-style-type: none"> 1. Los NIH deberían proporcionar recursos sobre cómo adaptar el compromiso público a las propuestas previstas para este tipo de compromisos de investigación y podrían exigir un plan para este compromiso público como parte de la propuesta. 2. Los compromisos públicos deben utilizarse para informar del diseño del ensayo y de cómo se proporciona la información a los participantes. Dada la novedad de presentar explicaciones generadas por IA a los participantes en ensayos clínicos, puede ser necesario revisar los procesos y la infraestructura pertinentes. La participación pública debe servir de base para tales cambios, con justificaciones claras si las opiniones del público no se incorporan finalmente a los programas y políticas.
<p>3. Los NIH, en consulta con otras agencias federales según proceda, deberían alinear sus políticas sobre datos y gobernanza de la investigación con valores públicos ampliamente extendidos.</p>	<ol style="list-style-type: none"> 1. Las reacciones del público pueden indicar un desajuste entre los valores públicos y las políticas actuales sobre la realización de ensayos clínicos, especialmente cuando se trata de resultados generados por IA. Por ejemplo, las conversaciones de la comunidad en este esfuerzo indicaron el apoyo público a la transparencia en el uso de modelos o algoritmos. Si se descubriera que ensayos clínicos como el del Estudio de Investigación carecen de transparencia, podrían ser necesarias nuevas políticas específicas sobre el uso de la IA en ensayos clínicos (o quizá sobre el uso de la IA en toda la investigación biomédica).
<p>4. Los NIH deben ayudar a establecer relaciones a largo plazo entre los miembros del público y las comunidades de investigación para facilitar la participación pública sostenida en las cuestiones en evolución relativas al uso, análisis, gestión y reutilización de los datos personales de salud en la investigación biomédica.</p>	<ol style="list-style-type: none"> 1. Como se describe en la Recomendación 2, los NIH deberían elaborar directrices y mejores prácticas sobre la participación pública que pudieran adaptarse fácilmente a los estudios de investigación propuestos y exigir un plan de participación pública como condición para la concesión. Cuando proceda, los NIH deben apoyar el desarrollo y mantenimiento de redes entre comunidades e investigadores. 2. El investigador y el equipo de investigación deben consultar las mejores prácticas de participación comunitaria e incorporarlas a su proceso de investigación. La participación de la comunidad en los usos novedosos de las LLM debe producirse antes de la conceptualización del estudio, informar el diseño del estudio y continuar después de que el estudio se haya completado. Los NIH deberían desempeñar un papel decisivo a la hora de facilitar este tipo de compromiso a largo plazo.

<p>5. Los NIH deberían animar a los investigadores a abordar de forma prospectiva en sus programas y propuestas de investigación las consideraciones bioéticas en áreas emergentes de la ciencia de datos que son particularmente importantes para las comunidades participantes, tal y como se refleja en el compromiso público.</p>	<ol style="list-style-type: none"> 1. Los NIH deben seguir apoyando los programas de bioética, los núcleos y los esfuerzos institucionales para integrar la bioética en la investigación biomédica. 2. El estudio de investigación y la propuesta deben incluir consultas con especialistas en bioética y expertos en participación pública. Las cuestiones pertinentes deben abordarse en el diseño del estudio, los planes de compromiso, la captación de participantes y los planes de difusión de datos y resultados al término del Estudio.
<p>6. Los NIH, en consulta con otras agencias federales según corresponda, deben desarrollar y difundir las mejores prácticas para informar a las comunidades participantes sobre cómo se gestionan, reutilizan y vinculan sus datos personales de salud con otros datos cuando el consentimiento u otra notificación de este tipo se extienden más allá de los requisitos reglamentarios.</p>	<ol style="list-style-type: none"> 2. La novedad de compartir los resultados generados por la IA puede plantear retos únicos a la hora de explicar a los posibles participantes en la investigación cómo se pueden gestionar, reutilizar y vincular sus datos con otros. El equipo de investigación debe consultar las mejores prácticas existentes para explicar las prácticas de datos tanto dentro del proceso de consentimiento como a través de otros canales. A medida que se amplía la base de conocimientos sobre la IA, los NIH deben ayudar a difundir las mejores prácticas actualizadas.
<p>7. Los NIH deben aprovechar la investigación previa sobre el retorno de los resultados y apoyar una mayor participación pública para informar sobre las mejores prácticas para el retorno de los resultados de la investigación biomédica utilizando la ciencia de datos emergente y para determinar las necesidades de infraestructura de implementación.</p>	<ol style="list-style-type: none"> 1. Los NIH deberían desarrollar mejores prácticas y puntos a considerar para el retorno de resultados a través de la recopilación y evaluación de los esfuerzos actuales (recomendación 1). A medida que se amplía la base de conocimientos sobre la devolución de resultados generados por IA, los NIH deben ayudar a difundir las mejores prácticas actualizadas. 2. Dado que el estudio de investigación busca informar sobre las mejores prácticas en materia de retorno de resultados, el equipo de investigación tendría que incorporar ciertas mejores prácticas en su evaluación de las explicaciones generadas por la LLM. Por ejemplo, puede haber algunos resultados que sería perjudicial comunicar a los participantes, y probablemente habría que destacar los riesgos que conllevan en el proceso de consentimiento, siguiendo las mejores prácticas existentes.
<p>8. Los NIH deben apoyar la exploración de métodos para proporcionar beneficios y retorno de valor más allá de la devolución de los resultados de la investigación a los participantes y sus comunidades, cuando sea apropiado y coherente con la normativa vigente.</p>	<ol style="list-style-type: none"> 2. El equipo de investigación debe considerar formas alternativas de devolver valor a los participantes, más allá de proporcionar los resultados de su tomografía computarizada. Si los investigadores se benefician económicamente del estudio, o si los participantes proporcionan los datos utilizados para generar los resultados de la LLM, puede ser especialmente necesario ofrecer una compensación adicional. Los métodos alternativos que se apliquen deben basarse en pruebas empíricas obtenidas mediante la participación de la comunidad.

Como muestra la tabla, las recomendaciones del NExTRAC a los NIH son flexibles y pueden adaptarse fácilmente a la investigación que proponga el uso de ciencia de datos novedosa y emergente para aplicaciones biomédicas, incluido el uso de LLM de IA generativa. Además, este ejercicio refuerza cuántos recursos son necesarios para ayudar a los investigadores a integrar la ética y el compromiso público en sus programas de investigación. Siguiendo estas Recomendaciones, los NIH no solo cambiarían el panorama de cómo se integra la ciencia de datos en la investigación apoyada por los NIH, sino que podrían cultivar una comunidad de práctica estándar para liderar con el ejemplo.