

## GUÍA PARA PREPARAR BASES DE DATOS CLÍNICOS BRUTOS PARA PUBLICAR EN REPOSITARIOS PÚBLICOS

### Introducción

Compartir el conjunto de datos utilizado para llegar a las conclusiones extraídas en un artículo, así como cualquier dato adicional necesario para reproducir los resultados del estudio en su totalidad, puede ser un requisito para la publicación en algunas revistas, como por ejemplo las del grupo de Public Library of Science (PLOS). Otras revistas de gran prestigio, como Nature and Science, también han establecido requisitos claros de intercambio de datos. Esto está generando a los investigadores dudas que plantean a los comités de ética de la investigación.

### Criterios generales

- Los participantes en el estudio fueron informados en el consentimiento informado de la posibilidad de publicar sus datos de modo confidencial.
- Compartir datos es una buena práctica científica, que debería promoverse y facilitarse.
- Compartir datos debe hacerse de un modo que preserve la confidencialidad de los datos.
- Los datos brutos que compartir han de limitarse al mínimo necesario para reproducir los resultados.
- Si no fuera posible publicar los resultados de una investigación sin identificar a la persona que participó en la misma o que aportó muestras biológicas, tales resultados solo podrán ser publicados cuando haya mediado el consentimiento previo y expreso de aquélla.

### Definición de conjunto de datos

A los efectos de esta guía, el **conjunto de datos es la colección agregada de observaciones** de pacientes (incluida la información sociodemográfica y clínica) utilizada con el fin de producir el resumen de los hallazgos estadísticos presentados en el informe principal del proyecto de investigación, ya sea que se haya publicado previamente o no. El conjunto de datos se refiere al que contiene el nivel mínimo de detalle necesario para reproducir todos los números reportados en el documento.

### Procedimientos para compartir las bases de datos

**a. Consentimiento informado.** Ante todo, hay que comprobar si los sujetos participantes en el estudio fueron informados en el consentimiento informado para publicar sus datos de modo confidencial. Si no lo hicieron, habrá que saber las razones de ello (la más habitual es que esta cuestión no se les planteó cuando se les pidió el consentimiento para el estudio; se debería promover que los participantes otorguen su consentimiento a la publicación de sus datos si es de modo confidencial).

**b. Comprobación del tipo de variables que compartir.** El procedimiento a seguir puede variar atendiendo al tipo de variables que se quiere compartir, ya sea **variables identificadoras directas** o **indirectas**. (Ver tabla más abajo).

Variables identificadoras directas: en ningún caso deberán aparecer en las bases de datos que compartir.

Variables identificadoras indirectas: el riesgo de reidentificación de los sujetos aumenta si la base de datos contiene tres o más de estas variables. En esos casos, el investigador debería consultar siempre al comité de ética de la investigación antes de compartir los datos.

**c. Dictamen del Comité.** Después del examen de cada caso, el comité debería pronunciarse sobre

- si la base de datos presenta un riesgo despreciable para la confidencialidad,
- si hay razones poderosas para compartir datos que superen los riesgos para la confidencialidad o
- si el riesgo de confidencialidad supera las ventajas de compartir los datos y requiere políticas más restrictivas de compartir los datos.

En todos los casos, el comité de ética de la investigación deberá asumir la responsabilidad de su dictamen ante los investigadores y las revistas científicas que piden compartir los datos.

**d. Medidas alternativas.** Se deben buscar técnicas que permitan compartir los datos siempre que sea posible, de modo que solo en casos excepcionales se deniegue el compartir los datos. Entre esas técnicas, sin ánimo de exhaustividad están:

- modificar fechas de modo uniforme en todos los casos (por ejemplo, añadiendo en todos los casos una misma cantidad de tiempo, de modo que el análisis estadístico no se altera).
- codificar variables geográficas que impidan su identificación.

<b>Variables identificadoras directas</b>	<b>Variables identificadoras indirectas</b> (el riesgo de reidentificación de los sujetos aumenta si la base de datos contiene tres o más de estas variables)
Nombre, Iniciales, dirección, incluido el código postal Números de teléfono, fax, información de contacto. Direcciones de correo electrónico Nº de identificación únicos (DNI, Nº seg. Social, identificadores de vehículos...) Identificadores de dispositivos médicos Direcciones web (URL) o de protocolo de Internet (IP) Datos biométricos (huellas dactilares...) Fotografía facial o imagen comparable. Cintas de audio Nombres de familiares Fechas relacionadas con una persona (incluida la fecha de nacimiento, admisión y alta hospitalaria, fallecimiento)	Edad*—o año de nacimiento—, sexo Diagnóstico, tratamiento recibido, centro de salud en donde se trata** Lugar de nacimiento, datos socioeconómicos — ocupación “rara” o lugar de trabajo, ingresos, nivel educativo— Composición familiar, medidas antropométricas, embarazos múltiples Etnia, denominadores poblacionales pequeños —de menos de 100 personas—, numeradores o cifras de recuento pequeños —menos de 3— Respuestas literales o transcripciones  *La edad es potencialmente identificable si el período de reclutamiento es corto y está completamente descrito. **Podría deducirse de las afiliaciones de los investigadores.

Fuente: Hrynaskiewicz I, Norton ML, Vickers AJ, Altman DG. Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. *BMJ* 2010;340c181.