



Cómo hacer investigación en Nutrición Clínica y no morir en el intento

Trucos para construir la base de datos

Tips for building the database

Vicente Lustres Pérez

CEO Biostattech S.L., Biostattech, Advice, Training & Innovation in Biostatistics, S.L.



El proceso inicial de cualquier investigación se asienta en cuatro pasos esenciales:

1. Tener la pregunta a investigar.
2. Determinar el/los objetivo/s de estudio (contestar a esa pregunta).
3. Plantear la hipótesis (es decir, formular la potencial respuesta a la pregunta planteada).

4. Diseño del estudio (lo que supone establecer la metodología que permite contestar a la pregunta).

En el planteamiento metodológico de un estudio se deben considerar numerosos aspectos: el diseño del estudio (tipo de estudio), los sujetos (población) de estudio, las variables (determinaciones principales, tanto las independientes -las exposiciones- como las dependientes -los efectos-), el modo de recogida y análisis de datos, y el cumplimiento de las normas éticas y legales.

TRUCOS

Dentro de la planificación y recopilación de datos, es fundamental cumplir tres requisitos: desarrollar la hipótesis principal y definir las medidas de resultados (*outcomes*); ser exactos y concretos al definir las variables; y determinar el tipo de dato más adecuado para cada variable. Atendiendo a estas exigencias, se debe evitar que la base de datos a desarrollar tenga demasiadas variables y datos, recogiendo principalmente aquellos que realmente sean relevantes para la investigación; por otro lado, resulta crucial incluir variables esenciales, pero tampoco simplificar en exceso la investigación; finalmente, es fundamental no perder información de partida.

Las variables de estudio pueden ser cualitativas o cuantitativas. Dentro del primer grupo, pueden dicotómicas (sí/no), nominales (color de ojos, por ejemplo) u ordinales (nivel de gravedad de un evento adverso, por ejemplo). Las cuantitativas pueden ser discretas (número de hijas/os, por ejemplo) o continuas (altura).

Para la correcta estandarización de variables, algunos consejos prácticos:

Conflicto de intereses: el autor declara no tener conflicto de interés.

Inteligencia artificial: el autor declara no haber usado inteligencia artificial (IA) ni ninguna herramienta que use IA para la redacción del artículo.

Lustres Pérez V. Trucos para construir la base de datos. Nutr Hosp 2025;42(N.º Extra 2):10-11

DOI: <http://dx.doi.org/10.20960/nh.06347>

©Copyright 2025 SENPE y ©Arán Ediciones S.L. Este es un artículo Open Access bajo la licencia CC BY-NC-SA (<http://creativecommons.org/licenses/by-nc-sa/4.0/>).

1. Codifica las variables (especialmente las categóricas), ya que esto facilita el análisis posterior de los datos.
2. Incluye siempre variables (*outcomes*, *covariables*) confusoras (como los datos demográficos).
3. Recolecta y almacena los datos clave, no solo las codificaciones o estimaciones.
4. Evita datos faltantes y codifica los *missings*, teniendo en cuenta que 0 es 0 y no un valor perdido.

En cuanto a la recolección de datos, es importante minimizar las entradas de texto libre en estudios cuantitativos, crear protocolos de procedimientos para el estudio (librería de códigos), asegurar el correcto funcionamiento del formulario para recolectar los datos, y efectuar una adecuada gestión y control de los datos sensibles.

Dentro de la amplia variedad de recursos y herramientas para crear la base de datos de un proyecto de investigación, resalta la plataforma REDCap (*Research Electronic Data Capture*), una herramienta segura y de uso sencillo para la recolección, gestión y análisis de datos en investigaciones científicas, especialmente en el ámbito clínico y académico. Permite la captura de datos electrónicos y ofrece una metodología de flujo de trabajo para diseñar bases de datos de investigación de ensayos clínicos e investigación traslacional. En cualquier caso, y a pesar de ser un recurso riguroso y fiable, el resultado que se obtenga de la base de datos que se genere se fundamenta en que el diseño del estudio haya sido óptimo. Junto a este recurso, uno de los más empleados para la construcción

de una base de datos, es el Excel; hasta el 80 % de los datos que manejamos en bases de datos proceden de un documento hecho con este programa informático.

BUENAS PRÁCTICAS

Como recomendaciones fundamentales de buenas prácticas en la construcción de una base de datos con Excel (para evitar errores y facilitar posibles soluciones), es esencial:

1. Anonimizar los datos.
2. Definir el formato de celda en función del registro que se quiere incorporar:
 - Cualitativas: codificación bien definida; definir y listar categorías.
 - Cuantitativas: número -clarificar cuántos decimales se van a coger, si se separarán por puntos o por comas...; fechas: mismo formato).
3. Validar los datos, instaurando reglas para evitar la introducción de registros erróneos (Fig. 1).

Ya para optimizar tiempos, es fundamental automatizar los cálculos mediante fórmulas, diseñar alertas para identificar errores, utilizar formularios estructurados (el identificador [ID] es clave para enlazar y fusionar todos los datos) y emplear cuestionarios validados.

Y que no se olvide que hasta el 30 % del tiempo que dedica un estadístico al analizar una base de datos se dedica a la depuración de datos, de ahí la importancia de que estos se introduzcan de la forma más correcta y uniforme posible.

Trucos para construir la base de datos

Buenas prácticas (errores y soluciones)

ID	Evento	Sexo	País	Fecha	Peso	Estadio
1	0	1	España	01/02/1800	75	N0
2	1	2	Italia	03/12/2020	48.5	N2a
3	2	2	Spain	4-12	480	#N/D
4	N	2	Italy	05/12/2021	59.8	N1
5	n	999		6-dic.		N0
6	no	h		44172	0	N0
7	No	homb		5/1282022	50-60	n0
8	s	hombre	spain	uno en 2019 y otro en 2020	50-70	NX
9	si	Hombre				N0
10	Si	Varón				N0
11	Si	v		08/12/2020		N0
12		m	españa	12-03-18		N2b
13	si_duplicated_si	muj	si_duplicated_si	2018-03-13		N0
14	si_duplicated_si	mujer	marzo-18			nX
15	si_duplicated_si	Mujer		15/03/2019		N1
16	si_duplicated_si	f				N2C
17	si_duplicated_si			26/06/2018		n0

Buenas prácticas (optimizar tiempos)

ID	Fecha de nacimiento	Fecha diagnóstico	Fecha inicio tratamiento	Fecha actualización	Edad diagnóstico	Edad inicio tratamiento	Edad actual
1	26/02/1971	25/02/2021	10/04/2024	31/03/2025	50,0	53,2	54,1
2	05/08/1980	14/09/2022	09/03/2024	31/03/2025	42,1	43,6	44,7
3	04/09/1991	05/04/2020	08/05/2022	31/03/2025	28,6	30,7	33,6
4	05/07/1965	15/10/2021	10/01/2023	31/03/2025	56,3	57,6	59,8
5	08/11/1999	14/08/2019		31/03/2025	19,8	-99,9	25,4

Figura 1.

Ejemplo de buenas prácticas en la construcción de bases de datos. En la parte izquierda se remarcan en verde ejemplos de registros correctos (atendiendo al criterio que previamente sea definido por los investigadores). En la parte de la derecha se muestran ejemplos de automatización. En rojo sería un registro incorrecto.