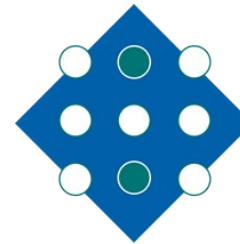




G CONSELLERIA
O SALUT I CONSUM
I SERVEI SALUT
B ILLES BALEARS



Institut
d'Investigació Sanitària
Illes Balears

IdISBa

Recerca amb bases de dades sanitàries

Plataforma de Recerca amb Informació Sanitària Illes Balears
(PRISIB)

Pau Pericàs

HUSE, 10 d'octubre de 2022

Pregunta

Protocol

PGD

Model
de dades

Extracció

Validació i
verificació

Anàlisi

Publicació

HEY, LOOK, WE HAVE A BUNCH
OF DATA! I'M GONNA ANALYZE IT.

NO, YOU FOOL! *THAT WILL
ONLY CREATE MORE DATA!*



Estudis observacionals amb RWD



Punts febles de l'assaig clínic:

- Evidència parcial
- Context limitat diferent a la pràctica habitual
- Resultats diferits
- Poc actualitzats
- Pacients no reals
- Elevat cost



Yetley, Elizabeth et Al. Options for basing Dietary Reference Intakes (DRIs) on chronic disease endpoints: report from a joint US-/Canadian-sponsored working group. American Journal of Clinical Nutrition, 2016.

Estudis observacionals amb RWD



Preguntes de recerca que l'assaig clínic no pot respondre i que requereixen de RWD:

- Com afecta l'aparició de noves tecnologies
- Malalties rares
- Identificació de variables rellevants per la presa de decisions
- Monitorització d'accés i ús de recursos sanitaris
- Detecció de necessitats de salut no satisfetes
- Detecció d'ús inadequat de recursos
- Impacte d'una intervenció en els processos assistencials

Estudis observacionals amb RWD



Avantatges de les RWD sobre els RCT:

- Es mesuren múltiples intervencions
- Sobre pacients reals
- En context real
- Amb evolució de les dades (Es poden ampliar les variables i fer-ne seguiment)
- Amb capacitat d'informar a temps dels canvis en els resultats

Limitacions i cauteles en l'ús de RWD:

- Dades incompletes (dificultat per saber el motiu de la incompletitud, mecanismes d'imputació)
- Dades no fidedignes (verificació, validació)
- La pròpia quantitat de dades (associacions espúries, anàlisis d'heterogeneïtats, multiplicitat de dimensions)
- Biaixos i factors de confusió (variança no explicada)

Estudis observacionals amb RWD

Pregunta

Protocol

PGD

Model
de dades

Extracció

Validació i
verificació

Anàlisi

Publicació

Al disseny de RCT es posa l'èmfasi en mesurar únicament l'efecte de la intervenció.

Amb RWD per aspirar a inferir una causalitat cal tenir un model causal que expliqui la relació causa - efecte que esperam observar i mitigar els biaixos i controlar els factors de confusió que puguin existir

Directed Acyclic Graphs

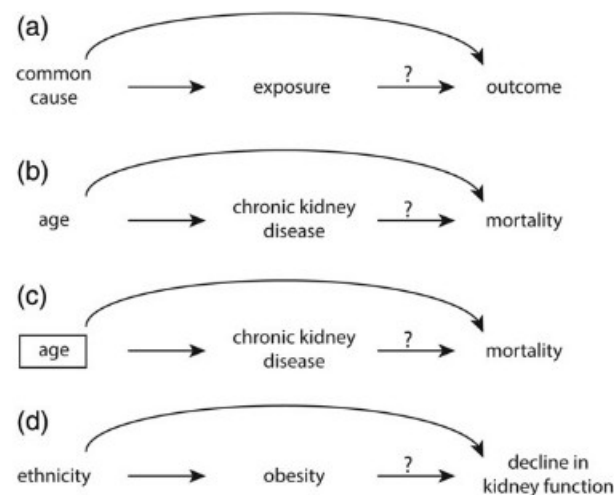


FIGURE 1: A graphical presentation of confounding in DAGs. (a) The structure of confounding in DAGs. Since age is a common cause of CKD and mortality, confounding is present when we want to assess the causal relationship between the exposure CKD and the outcome mortality (b). The backdoor path from CKD via age to mortality can be blocked by conditioning on age, as depicted by a box around age in (c). Similarly, ethnicity is a common cause of obesity and decline in kidney function (d). The backdoor path from obesity via ethnicity to decline in kidney function can be blocked by conditioning on ethnicity. If ethnicity is not measured or not properly measured, residual confounding remains present.

Suttorp, M. M et Al. Graphical presentation of confounding in directed acyclic graphs. 2015 Nephrology Dialysis Transplantation

Estudis observacionals amb RWD

Pregunta

Protocol

PGD

Model
de dades

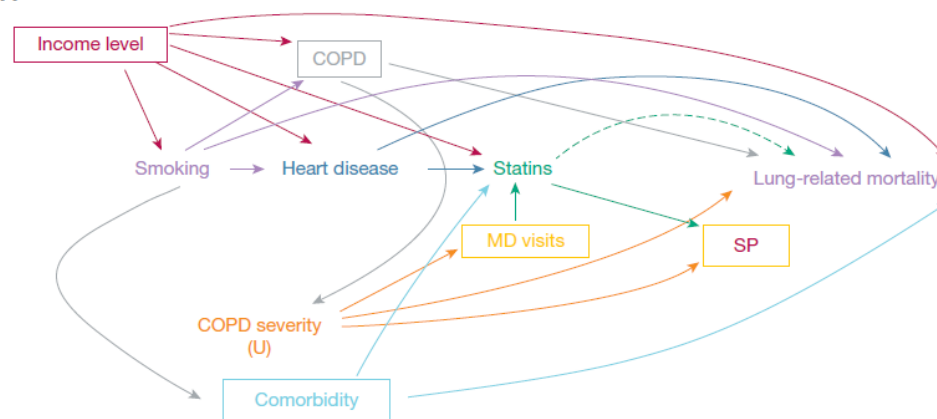
Extracció

Validació i
verificació

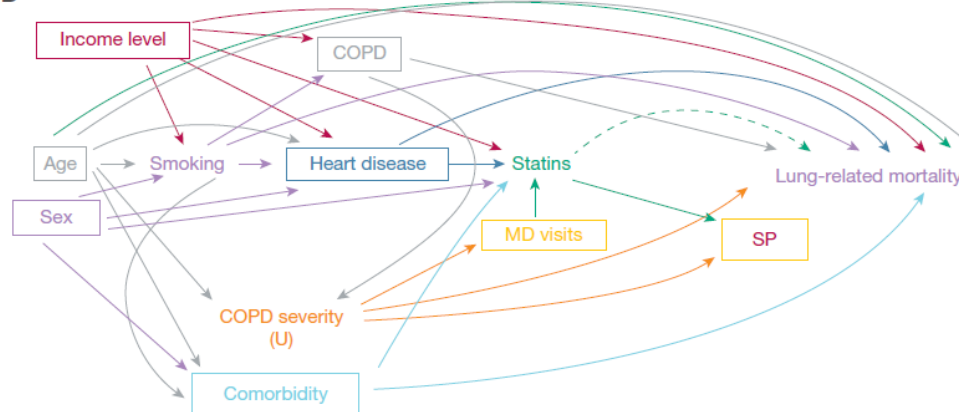
Anàlisi

Publicació

A

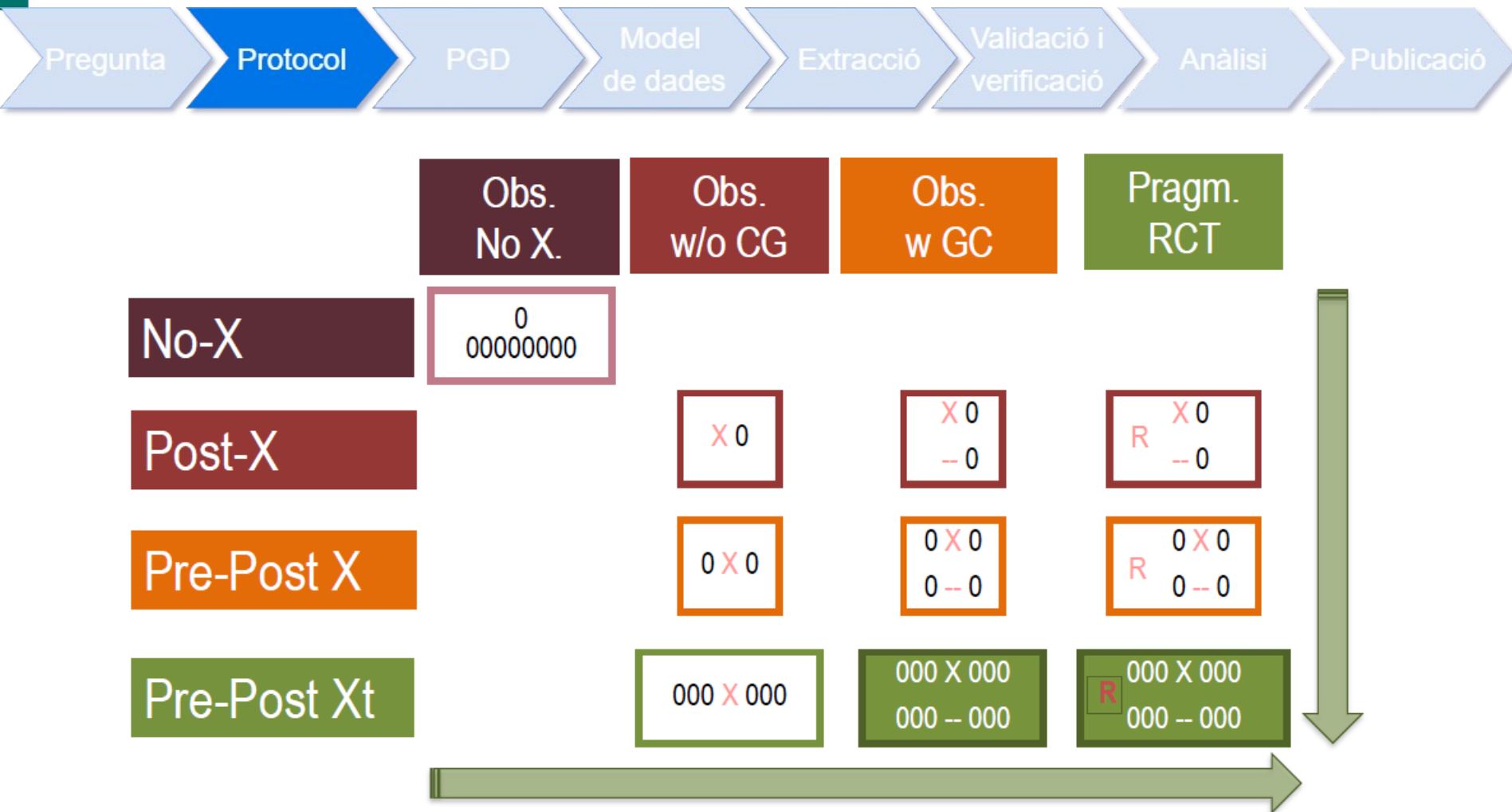


B



Etminan M. Using Causal Diagrams to Improve the Design and Interpretation of Medical Research. Chest. 2020

Estudis observacionals amb RWD



Enrique Bernal, Francisco Estupiñán. C5 Utilización de datos masivos de vida real en la investigación en servicios y políticas sanitarias, XXXIII Escola de Salut Pública de Menorca

Estudis observacionals amb RWD

Pregunta

Protocol

PGD

Model
de dades

Extracció

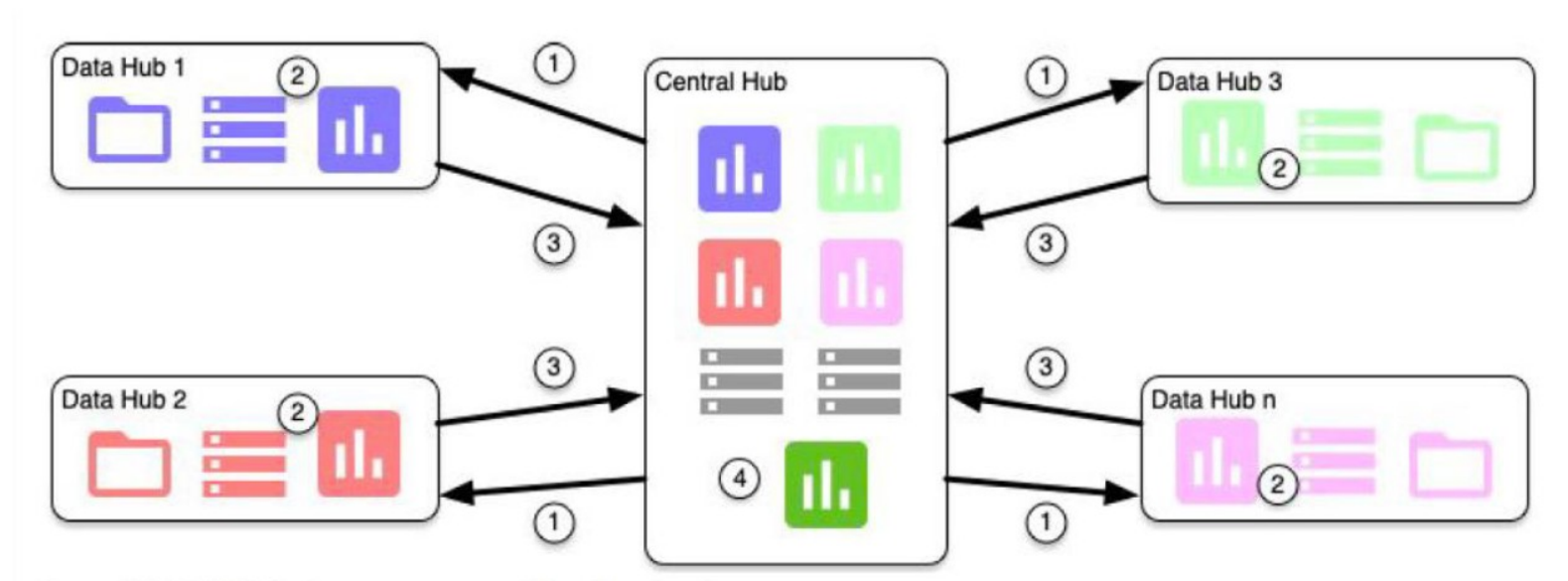
Validació i
verificació

Anàlisi

Publicació

Model federat

Compartir mètodes i resultats sense donar accés a les dades



<https://www.phiri.eu/>

Estudis observacionals amb RWD



Pla de Gestió de Dades

És un document formal que s'ha de presentar a l'inici del projecte, en el qual se esbossa la gestió que s'aplicarà durant el cicle de vida de les dades, des de la seva creació fins més enllà de la finalització del projecte

- 01** Revisa los **requerimientos** de la entidad financiadora.
- 02** **Identifica los datos:** tipología, procedencia, volumen, formatos y ficheros.
- 03** **Define cómo se organizarán y gestionarán los datos:** nombre de los ficheros, control de versiones, software necesario...
- 04** **Explica cómo se documentarán los datos:** identifica la información a procesar, consulta si hay estándares o esquemas de metadatos, identifica las herramientas que permitan gestionarlos.
- 05** Describe los procesos que aseguran una **buena calidad de los datos**.
- 06** **Prepara una estrategia de almacenamiento** (durante el proceso) y de **preservación de datos** (repositorio).
- 07** **Define las políticas de datos del proyecto:** cuestiones sobre propiedad intelectual y cómo se tratarán los datos sensibles y personales.
- 08** **Describe cómo se difundirán los datos:** dónde, cuáles, cuándo se van a difundir. Si publicas los datos en un repositorio, como información suplementaria del artículo o como *data paper*.
- 09** **Asigna roles y responsabilidades** para las personas y organizaciones participantes del proyecto.
- 10** **Prepara un presupuesto realista:** la gestión de datos cuesta tiempo y dinero en términos de software, hardware, servicios y personal.

Maseda, Irene. 10 pasos para elaborar un Plan de Gestión de Datos, 2020, CNIC-ISCIII

Estudis observacionals amb RWD



Hauria d'explicar com s'aplicaran els principis FAIR



Findable: Metadata and data should be easy to find for both humans and computers.



Accessible: The exact conditions under which the data is accessible should be provided in such a way that humans and machines can understand them.



Interoperable: The (meta)data should be based on standardized vocabularies, ontologies, thesauri etc. so that it integrates with existing applications or workflows.



Reusable: Metadata and data should be well-described so that they can be replicated and/or combined in different research settings.

Estudis observacionals amb RWD



I The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards

Wilkinson, M. D., et Al. Comment: The FAIR Guiding Principles for scientific data management and stewardship. 2016. Scientific Data

Estudis observacionals amb RWD



"ALL RESEARCH SHOULD AIM
TO BE F.A.I.R."

#FIGSHAREFEST

	GOOD	BAD
FINDABLE	ONLINE DATABASE	FILING CABINET IN A BATH IN THE BASEMENT UNDER A LEAKING PIPE
ACCESSABLE	OPEN ACCESS FOR EVERYONE (NO LOGIN)	THE FILING CABINET ALSO IS HOME TO A NEST OF WILD BADGERS
INTEROPERABLE	ALL DATA IS IN OPEN FORMATS	ALL DOCUMENTS ARE PRINTED IN COMIC SANS AND WRITTEN IN ESPERANTO
REUSEABLE	GOOD META DATA AND SECURELY STORED FOR 10 YEARS	THE PAPER EXPLODES IF IT'S READ

ERRANTSCIENCE.COM

Estudis observacionals amb RWD



Metadata

Documentació detallada de les dades que es volen obtenir redactat entre l'equip de recerca i els tècnics que realitzen l'extracció

PRISIB - MODEL DE DADES DE PROJECTE DE RECERCA									
						Data	Versió	Autor	Canvis
						Històric de versions	06/10/22	1 Pau Pericàs	
Nom Projecte	HCE: Anàlisis del impacto de determinantes sociales y morbilidad en atención primaria sobre resultados de salud poblacionales combinando grandes bases de datos								
Codi Projecte	P2101								
Investigador Principal	Ana Clavería								
Cas d'ús	<ul style="list-style-type: none">• Predecir resultados en salud medidos como morbilidad hospitalaria considerando como predictor a la morbilidad de atención primaria ajustada por determinantes sociales, combinando datos de la historia clínica electrónica y diversas encuestas del INE• Analizar variabilidad geográfica del impacto de la morbilidad de atención primaria ajustada por determinantes sociales, sobre resultados de salud como la morbilidad hospitalaria, combinando datos de la historia clínica electrónica y diversas encuestas del INE.								
Versió Model de dades	1								
Autor	Pau Pericàs, Ana Clavería, Beatriz González								

Estudis observacionals amb RWD



Identifica les dades, perquè es recullen, en quins formats, de quines fonts i amb quines codificacions

[illegible]

Estudis observacionals amb RWD

Pregunta

Protocol

PGD

Model
de dades

Extracció

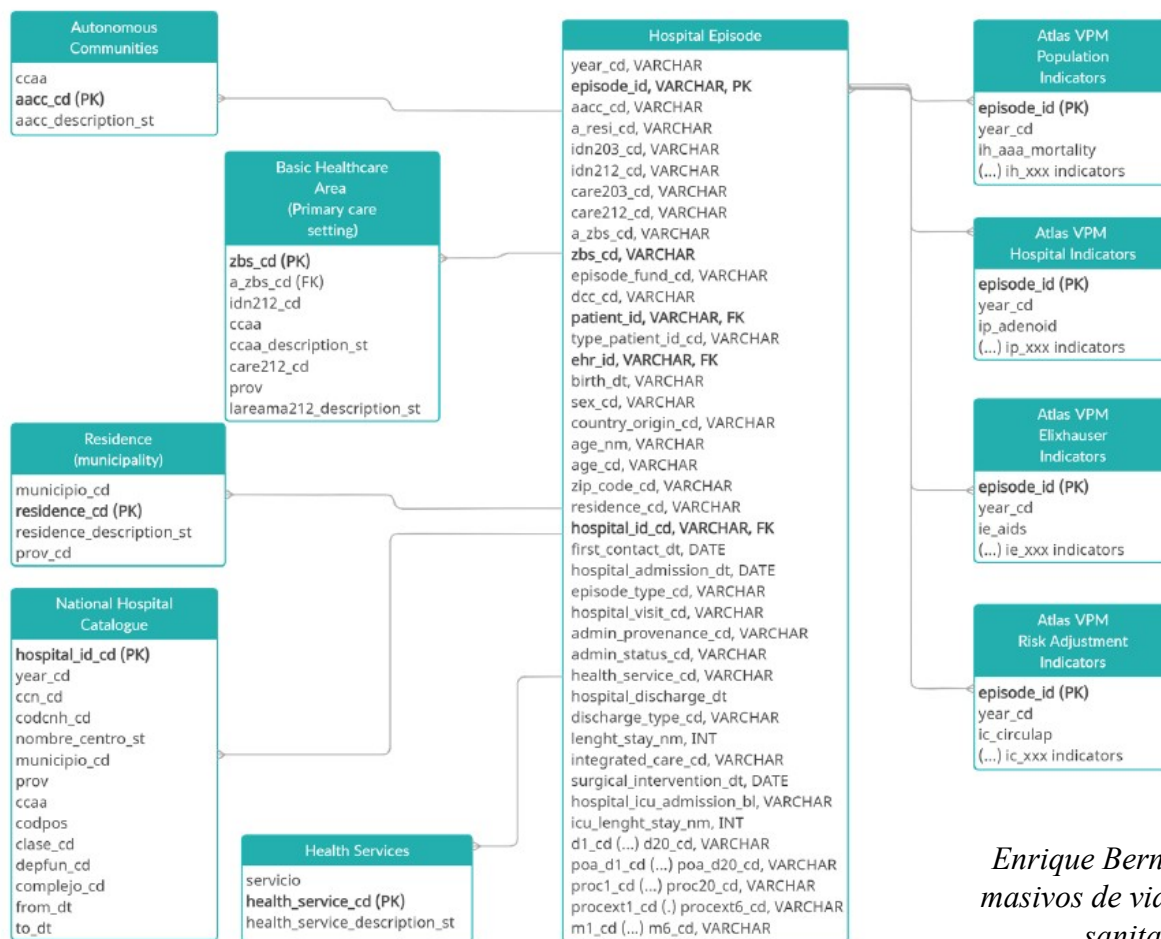
Validació i
verificació

Anàlisi

Publicació

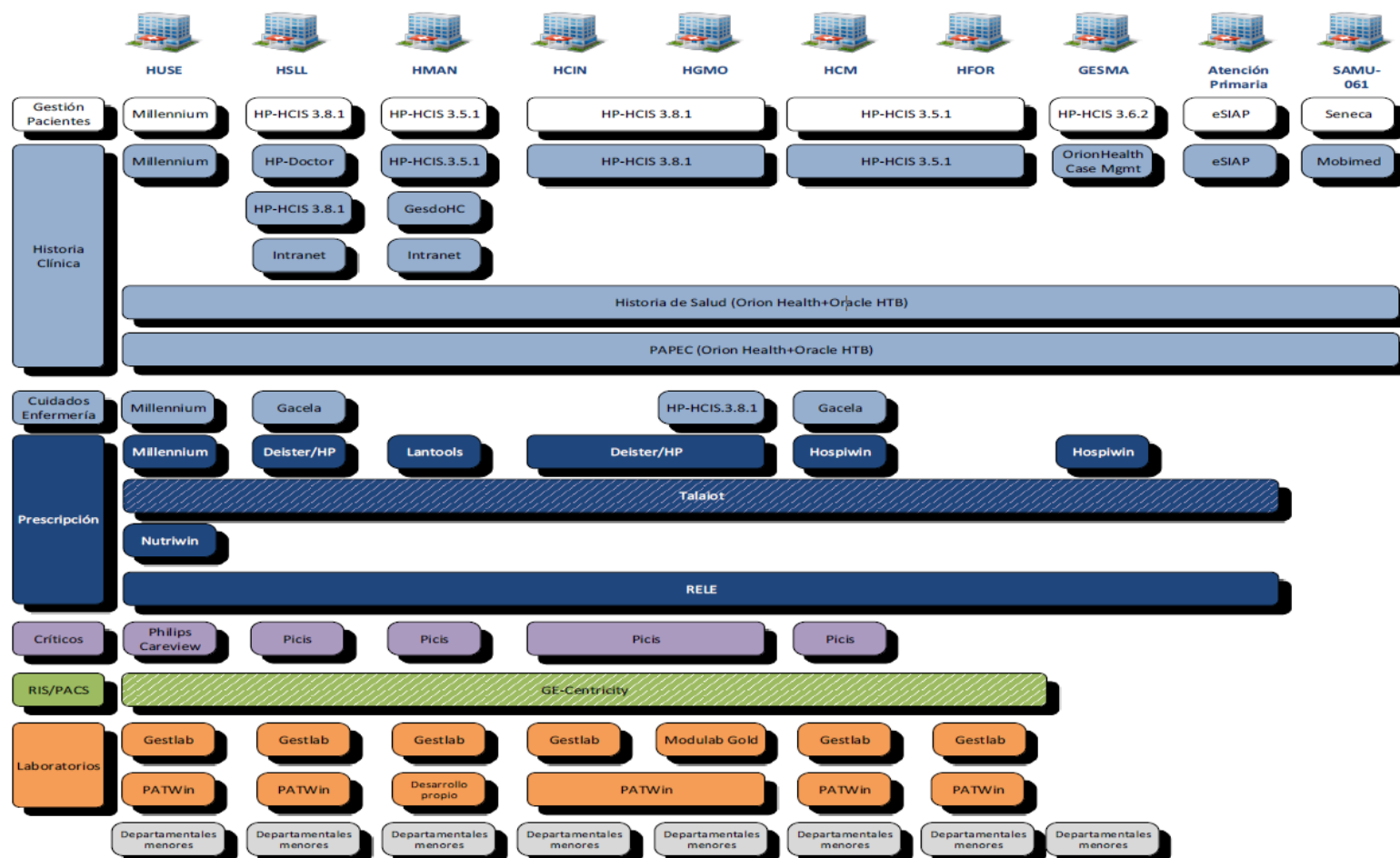
Entitat	Variable	Descripció	Tipus	Format	Unitat	Nivell de requeriment	Diccionari de classificació	Regla de validació	Observations	Possible font de dades	Possible taula d'origen
PACIENTE	ID_ESTUDIO	Identificador único del paciente	COMPUESTO	15					PK	eSIAP	FZM T ESI FUSUARIO
PACIENTE	ID_NODO	Código único del nodo	STR	Númerico (3)				d{3}		eSIAP	FZM T ESI FUSUARIO
PACIENTE	ID_PACIENTE	Código único dentro del nodo	STR	Númerico (12)				d{12}		eSIAP	FZM T ESI FUSUARIO
PACIENTE	SEXO	Sexo del paciente	CHAR	Carácter único, mayúsculas (1)			HL70001	[FMUJ], F=Mujer, M=Hombre, U=Des conocido		eSIAP	FZM T ESI FUSUARIO
PACIENTE	FECHA DE NACIMIENTO	Fecha de nacimiento	DATE	YYYYMMDD				d{8}, 19000101-20190101		eSIAP	FZM T ESI FUSUARIO
PACIENTE	FECHA DE EXITUS	Fecha del exitus	DATE	YYYYMMDD				d{8}, 19000101-20190101		eSIAP	FZM T ESI FUSUARIO
PACIENTE	NACIONALIDAD	Nacionalidad	COMPUESTO	43			HL70399		códigos ISO	eSIAP	FZM T ESI FUSUARIO
PACIENTE	CÓDIGO PAIS	Código del país	STR	Texto (3)			HL70399	[A-Z]{3}		eSIAP	FZM T ESI FUSUARIO
PACIENTE	NOMBRE PAIS	Nombre oficial	STR	Texto (40)			HL70399	{40}		eSIAP	FZM T ESI FUSUARIO
PACIENTE	COMUNIDAD AUTÓNOMA	ID de Comunidad autónoma de residencia	COMPUESTO	32					https://www.ine.es/daco/daco42/codmun/cod_ccaa.htm	eSIAP	FZM T ESI FUSUARIO
PACIENTE	CÓDIGO CCAA	Código de la CCAA de residencia	STR	Númerico (2)			Diccionario INE 2022	d{2}, '01'-19'		eSIAP	FZM T ESI FUSUARIO
PACIENTE	NOMBRE CCAA	Nombre oficial	STR	Texto (30)			Diccionario INE 2022	{30}		eSIAP	FZM T ESI FUSUARIO
PACIENTE	PROVINCIA	ID de Provincia de residencia	COMPUESTO	32					daco42/codmun/cod_provincia.htm	eSIAP	FZM T ESI FUSUARIO
PACIENTE	CÓDIGO PROVINCIA	Código de la provincia de residencia	STR	Númerico (2)			Diccionario INE 2022	d{2}, '01'-52'		eSIAP	FZM T ESI FUSUARIO
PACIENTE	NOMBRE PROVINCIA	Nombre oficial	STR	Texto (30)			Diccionario INE 2022	{30}		eSIAP	FZM T ESI FUSUARIO
PACIENTE	MUNICIPIO	ID de Municipio de residencia	COMPUESTO	55					daco42/codmun/22codmun.xlsx	eSIAP	FZM T ESI FUSUARIO
PACIENTE	CÓDIGO PROVINCIA	Código de la provincia de residencia	STR	Númerico (2)			Diccionario INE 2022	d{2}, '01'-52'		eSIAP	FZM T ESI FUSUARIO
PACIENTE	CÓDIGO MUNICIPIO	Código del municipio de residencia	STR	Númerico (3)			Diccionario INE 2022	d{3}, '001'-902'		eSIAP	FZM T ESI FUSUARIO
PACIENTE	NOMBRE MUNICIPIO	Nombre oficial	STR	Texto (50)			Diccionario INE 2022	{50}		eSIAP	FZM T ESI FUSUARIO
PACIENTE	ZONA BÁSICA DE SALUD	Identificador de Zona Básica	COMPUESTO	57						eSIAP	FZM T ESI FUSUARIO
PACIENTE	ID_NODO	Código único del nodo	STR	Númerico (3)				d{3}		eSIAP	FZM T ESI FUSUARIO
PACIENTE	CÓDIGO ZBS	Código único dentro del nodo	STR	Númerico (4)				d{4}		eSIAP	FZM T ESI FUSUARIO
PACIENTE	NOMBRE ZBS	Nombre oficial	STR	Texto (50)			Decretos de ordenación sanitaria	{50}	http://www.boa.aragon.es/cgi-bin/EOA/BRSCGI?CMD=VEROBJ&MLKOB=1191576225858&type=pdf	eSIAP	FZM T ESI FUSUARIO
PACIENTE	DISTRITO ANONIMIZADO	ID Distrito	COMPUESTO	7						eSIAP	FZM T ESI FUSUARIO
PACIENTE	ID_NODO	Código único del nodo	STR					d{3}		eSIAP	FZM T ESI FUSUARIO
PACIENTE	CÓDIGO DISTRITO	Código único dentro del nodo	STR	Númerico (4)			INE	d{4}	secciones_Aragón 1450, Baleares 662	eSIAP	FZM T ESI FUSUARIO
PACIENTE	APORTACION TSI	Tipo de aportación farmacéutica	STR	Alfanumérico (7)				TSI J 1 0 2 1-6 1 TSI 001'-TSI 006'		eSIAP	FZM T ESI FUSUARIO
PACIENTE	TIPO USUARIO	Tipo de usuario	STR	Mayúsculas (3)				{3}		eSIAP	FZM T ESI FUSUARIO
MORBILIDAD AP	ID_MORBILIDAD AP	Único en el estudio	COMPUESTO	18					PK	eSIAP	FZM T ESI HDIAGNOS
MORBILIDAD AP	ID_NODO	Código único del nodo	STR	Númerico (3)				d{3}		eSIAP	FZM T ESI HDIAGNOS
MORBILIDAD AP	NUMERO DE DIAGNÓSTICO	Único dentro de cada nodo	STR	Númerico(15)				d{15}		eSIAP	FZM T ESI HDIAGNOS
MORBILIDAD AP	ID_ESTUDIO	Identificador único del paciente	COMPUESTO	Númerico(15)					FK_PACIENTE	eSIAP	FZM T ESI HDIAGNOS
MORBILIDAD AP	FECHA	Fecha de registro del diagnóstico	DATE	YYYYMMDD				d{8}, 19000101-20190101		eSIAP	FZM T ESI HDIAGNOS
MORBILIDAD AP	DIAGNOSTICO	ID del diagnóstico	COMPUESTO	101						eSIAP	FZM T ESI HDIAGNOS

Estudis observacionals amb RWD



Enrique Bernal, Francisco Estupiñán. C5 Utilización de datos masivos de vida real en la investigación en servicios y políticas sanitarias, XXXIII Escola de Salut Pública de Menorca

Estudis observacionals amb RWD



Estudis observacionals amb RWD

Pregunta

Protocol

PGD

Model
de dades

Extracció

Validació i
verificació

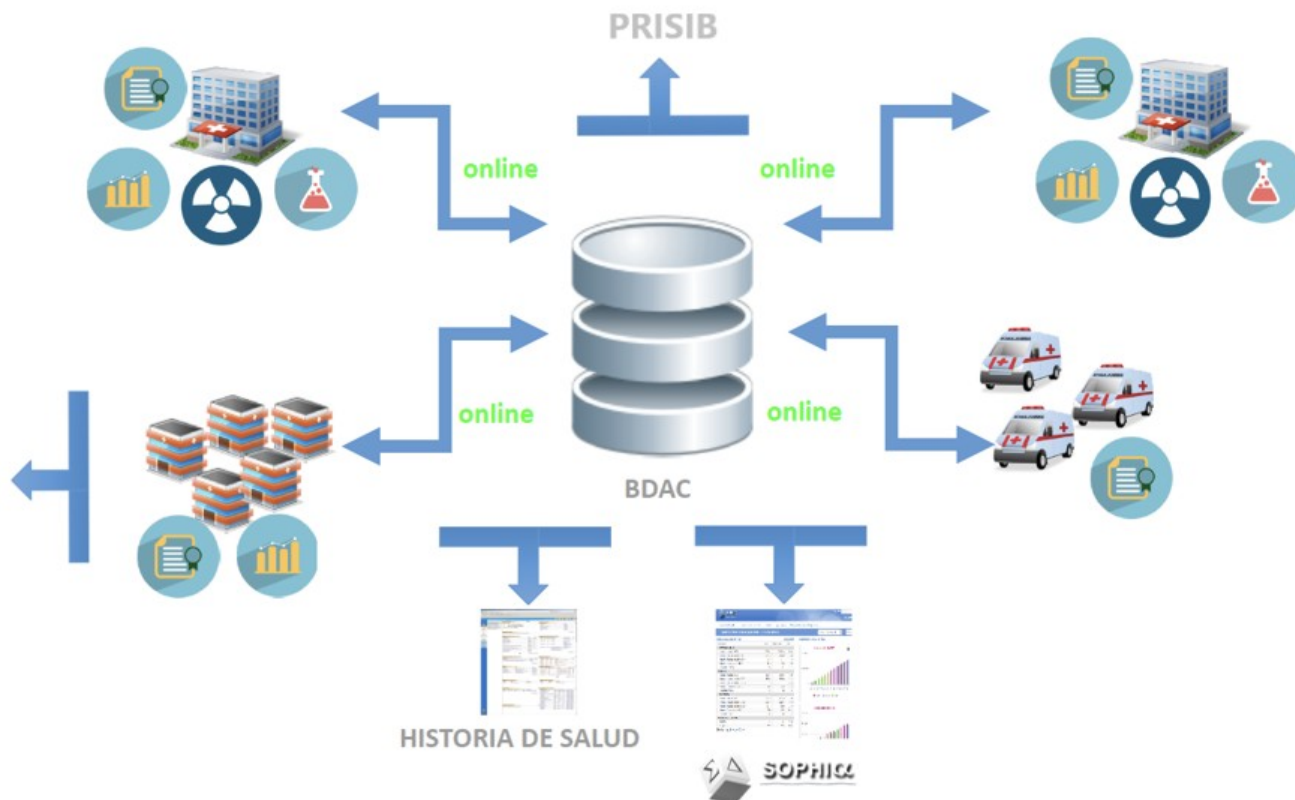
Anàlisi

Publicació

IAmetrics®



Institut
d'Investigació Sanitària
Illes Balears



Estudis observacionals amb RWD



Problemes:

- Múltiples models de dades
- Diferents sistemes de codificació
- Dades incompletes
- Dades esbiaixades
- Documentació escassa i desfasada
- Diferents necessitats

Estudis observacionals amb RWD



La qualitat de les dades depèn del seu propòsit

“Les dades són d'alta qualitat si són adequades pels usos prevists en les operacions, la presa de decisions i la planificació.

Les dades són aptes per el seu ús si no tenen defectes i tenen les característiques desitjades”

Data Quality: The Field Guide, Thomas C. Redman, Ph.D. Digital Press, 2001

Estudis observacionals amb RWD



Dimensions de la qualitat de les dades:

- Exactitud: Representen les dades correctament l'entitat o l'esdeveniment real?
- Consistència: Les dades no contenen contradiccions?
- Disponibilitat: Es pot accedir a les dades ara i al llarg del temps?
- Integritat: Inclouen les dades tots els elements que representen l'entitat o l'esdeveniment?
- Completitud: Inclouen les dades a tots els elements que es pretenen representar?
- Conformitat: Segueixen les dades les normes acceptades?
- Processabilitat: Són les dades llegibles per les màquines?
- Credibilitat: Es basen en fonts fiables?
- Pertinència: Inclouen una quantitat adequada de dades?
- Vigència: Representen la situació real i es publiquen amb prou antelació?

- Sidi,Fatimah, et al. "Data quality: A survey of data quality dimensions." 2012

Estudis observacionals amb RWD



Verificació - Comprovació de que les dades són correctes segons una sèrie de regles internes

Les dates tenen els format DD/MM/AAAA

La variable X està present a tots els registres

Els identificadors relacionen els diferents registres correctament

Validació - Comprovació de que les dades representen la realitat que es pretén mesurar segons referències externes

Una variable identifica els pacients amb una malaltia correctament

La distribució d'una variable coincideix amb l'esperable

Estudis observacionals amb RWD



VERIFICATION		VALIDATION	
DEFINITION	EXAMPLE	DEFINITION	EXAMPLE
CONFORMANCE: DO DATA VALUES ADHERE TO SPECIFIED STANDARDS AND FORMATS?			
VALUE CONFORMANCE			
a. Data values conform to internal formatting constraints.	a. Sex is only one ASCII character.	a. Data values conform to representational constraints based on external standards.	a. Values for primary language conform to ISO standards.
b. Data values conform to allowable values or ranges.	b. Sex only has values "M," "F," or "U."		
RELATIONAL CONFORMANCE			
a. Data values conform to relational constraints.	a. Patient medical record number links to other tables as required.	a. Data values conform to relational constraints based on external standards.	a. Data values conform to all not-NULL requirements in a common multi-institutional data exchange format.
b. Unique (key) data values are not duplicated.	b. A medical record number is assigned to a single patient.		
c. Changes to the data model or data model versioning.	c. Version 1 data does not include medical discharge hour.		
COMPUTATIONAL CONFORMANCE			
a. Computed values conform to computational or programming specifications.	a. Database- and hard-calculated Body Mass Index (BMI) values are identical.	a. Computed results based on published algorithms yield values that match validation values provided by external source.	a. Computed BMI percentiles yield identical values compared to test results and values provided by the CDC.

Kahn, M. G. Et al. Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data 2016

Estudis observacionals amb RWD



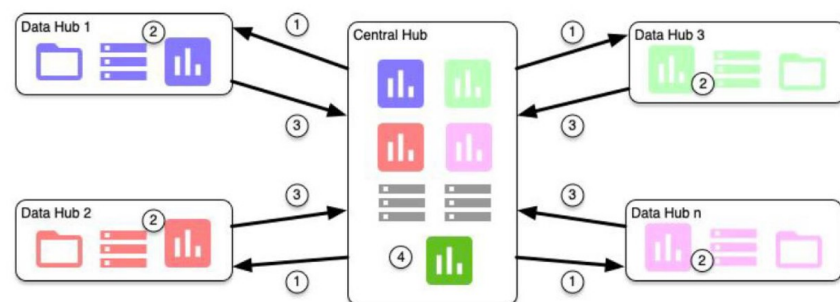
Estudis observacionals amb RWD



Data Quality Assessment

Result report

Aggregated Output



year	month	age_cd	sex_cd	confirmed_cases	hospital_admissions	icu_admissions
2021		115-19	0	391	37	1
2021		115-19	1	389	29	5
2021		120-24	0	406	29	2
2021		120-24	1	395	31	1
2021		125-29	0	405	38	3
2021		125-29	1	376	24	0
2021		130-34	0	369	31	1
2021		130-34	1	371	49	2
2021		135-39	0	421	45	4
2021		135-39	1	413	39	4
2021		140-44	0	399	29	2
2021		140-44	1	381	37	2
2021		145-49	0	419	42	0
2021		145-49	1	423	41	2
2021		150-54	0	396	39	5
2021		150-54	1	378	42	2

Javier González-Galindo, et Al. (2022). PHIRI APP - WP7 - PHIRI Federated Research Infrastructure (FRI) - Zenodo.

Estudis observacionals amb RWD



Confidentiality

Physicians are obligated **not to disclose confidential information given by a patient** to another party without the patient's authorization.

Autonomy

All persons have intrinsic and unconditional worth, and therefore, **should have the power to make rational decisions and moral choices**, and each should **be allowed to exercise** his or her capacity for **self-determination**

Beneficence

The principle of beneficence is the **obligation** of physician **to act for the benefit of the patient** and supports a number of moral rules to **protect and defend the right of others, prevent harm**, remove conditions that will cause harm, help persons with disabilities, and rescue persons in danger.

Estudis observacionals amb RWD



Anonimització

87% (216 million of 248 million) of the population in the United States had reported characteristics that likely made them unique based only on {5-digit ZIP, gender, date of birth}.

K-anonymity

Let $RT(A_1, \dots, A_n)$ be a table and QIRT be the quasi-identifier associated with it.

RT is said to satisfy k-anonymity if and only if each sequence of values in $RT[QIRT]$ appears with at least k occurrences in $RT[QIRT]$.

L. Sweeney. k-anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 2002

Estudis observacionals amb RWD

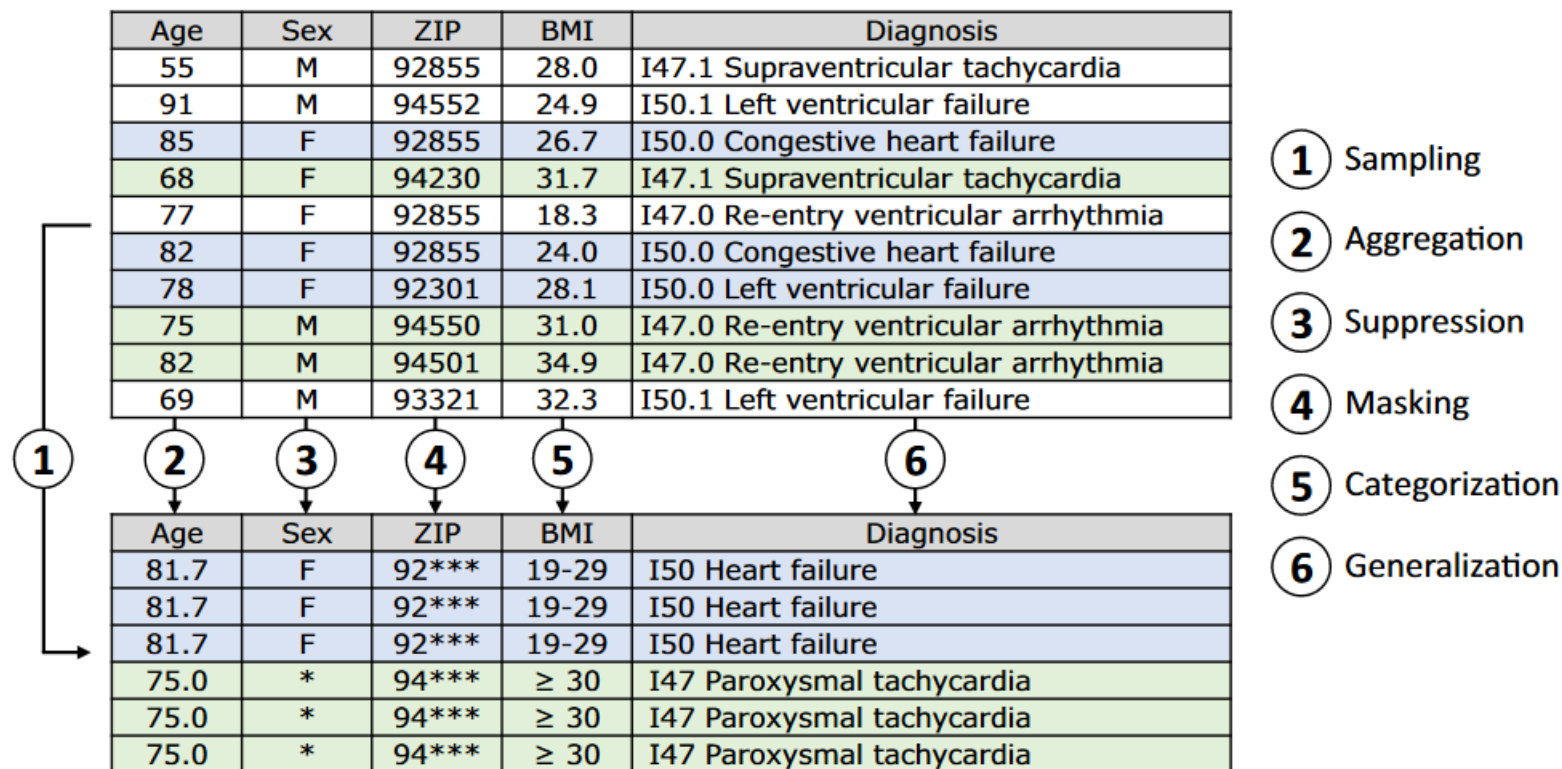


Figure 1: Exemplary anonymization process.

Estudis observacionals amb RWD



The OpenAIRE Consortium is a Horizon 2020 (FP8) project, aimed to support the implementation of the EC and ERC Open Access policies.

Its successor OpenAIREplus is aimed at linking the aggregated research publications to the accompanying research and **project information, datasets and author information.**



The Zenodo research data repository is a product of OpenAIRE.

Estudis observacionals amb RWD

Pregunta

Protocol

PGD

Model
de dades

Extracció

Validació i
verificació

Anàlisi

Publicació

zenodo

Search



Upload

Communities

Log in

Sign up

All versions

Found 6424 results.

< 1 2 3 4 5 6 7 8 9 >

Sort by:

Best match

asc

View

June 24, 2022 (1.0.1) Report Open Access

PHIRI - WP6 - Use Case B outputs from the local analyses (interactive reports)

Enrique Bernal-Delgado, Francisco Estupiñán-Romero, Javier González-Galindo, Natalia Martínez-Lizaga, Pascal Derycke, Nienke Schutte, Liesbet Van Eycken, Nancy Van Damme, Kris Henau, Ronan Lyons, Sarah Aldridge, Andrea Faragalli, Flavia Carle, Rosaria Gesuita, Luigi Palmieri, Jānis Misiņš, Jakov Vuković, Jelena Dimnjaković

The PHIRI Federated Research Infrastructure (FRI) is supported by a containerized reproducible solution for data analysis (i.e. use cases) to be deployed on-premises by each participant partner (a.k.a PHIRI-app) to generate a set of local results (outputs) to be further analysed and interpreted.

Uploaded on June 24, 2022

1 more version(s) exist for this record

Type

- ☐ Publication (1597725) +
- ☐ Image (774418) +
- ☐ Dataset (164404)
- ☐ Software (86000)
- ☐ Presentation (32905)
- ☐ Other (15003)
- ☐ Poster (14088)
- ☐ Video (6428)
- ☐ Lesson (3964)
- ☐ PhysicalObject (708)

Why use Zenodo?

- **Safe** — your research is stored safely for the future in CERN's Data Centre for as long as CERN exists.
- **Trusted** — built and operated by CERN and OpenAIRE to ensure that everyone can join in Open Science.
- **Citeable** — every upload is assigned a Digital Object Identifier (DOI), to make them citable and trackable.
- **No waiting time** — Uploads are made available online as soon as you hit publish, and your DOI is registered within seconds.
- **Open or closed** — Share e.g. anonymized clinical trial data with only medical professionals via our restricted access mode.
- **Versioning** — Easily update your dataset with our versioning feature.
- **GitHub integration** — Easily preserve your GitHub repository in Zenodo.
- **Usage statistics** — All uploads display standards compliant usage statistics

Estudis observacionals amb RWD

Pregunta

Protocol

PGD

Model
de dades

Extracció

Validació i
verificació

Anàlisi

Publicació

We believe this resolves
all remaining questions
on this topic. No further
research is needed.

References

1. [Illegible]
2. [Illegible]
3. [Illegible]
4. [Illegible]

JUST ONCE, I WANT TO SEE A RESEARCH
PAPER WITH THE GUTS TO END THIS WAY.

<https://xkcd.com/2268/>

Bases Jurídiques

LLEI 41/2002, de 14 de novembre, bàsica reguladora de l'autonomia del pacient i de drets i obligacions en matèria d'informació i documentació clínica. («BOE» 274, de 15-11-2002.)

Article 16. *Usos de la història clínica.*

3. L'accés a la **història clínica amb finalitats** judicials, epidemiològiques, de salut pública, **d'investigació** o de docència, **es regeix pel que disposen la Llei orgànica 15/1999***, de protecció de dades de caràcter personal, **i la Llei 14/1986, general de sanitat**, i altres normes d'aplicació en cada cas.

L'accés a la història clínica amb aquestes finalitats **obliga a preservar les dades d'identificació personal del pacient, separades de les de caràcter clinicoassistencial**, de manera que com a regla general quedi assegurat l'anonimat, llevat que el pacient mateix hagi donat el seu consentiment per no separar-les.

Bases Jurídiques

Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.

Disposición adicional decimoséptima. *Tratamientos de datos de salud.*

2. El tratamiento de datos en la investigación en salud se regirá por los siguientes criterios:

- a) **El interesado** o, en su caso, su representante legal **podrá otorgar el consentimiento para el uso de sus datos con fines de investigación** en salud y, en particular, la biomédica. Tales finalidades podrán abarcar categorías relacionadas con áreas generales vinculadas a una especialidad médica o investigadora.
- b) **Las autoridades sanitarias** e instituciones públicas con competencias en vigilancia de la salud pública **podrán llevar a cabo estudios científicos sin el consentimiento de los afectados en situaciones de excepcional relevancia** y gravedad para la salud pública.
- c) Se considerará lícita y compatible la **reutilización de datos personales** con fines de investigación en materia de salud y biomédica cuando, **habiéndose obtenido el consentimiento** para una finalidad concreta, **se utilicen los datos para finalidades** o áreas de investigación **relacionadas** con el área en la que se integrase científicamente el estudio inicial.

Bases Jurídiques

Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.

Disposición adicional decimoséptima. *Tratamientos de datos de salud.*

d) Se considera lícito el **uso de datos personales seudonimizados** con fines de investigación en salud y, en particular, biomédica.

1.º Una **separación** técnica y funcional **entre el equipo investigador y quienes realicen la seudonimización** y conserven la información que posibilite la reidentificación.

2.º Que los datos seudonimizados únicamente sean accesibles al equipo de investigación cuando:

i) Exista un **compromiso expreso de confidencialidad** y de no realizar ninguna actividad de reidentificación.

ii) Se adopten **medidas de seguridad específicas** para evitar la reidentificación y el acceso de terceros no autorizados.

Podrá procederse a la reidentificación de los datos en su origen, **cuando con motivo de una investigación** que utilice datos seudonimizados, **se aprecie la existencia de un peligro real** y concreto para la seguridad o salud de una persona o grupo de personas, **o una amenaza grave** para sus derechos o sea necesaria para garantizar una adecuada asistencia sanitaria.

Bases Jurídiques

Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.

Disposición adicional decimoséptima. *Tratamientos de datos de salud.*

f) Cuando conforme a lo previsto por el artículo 89 del Reglamento (UE) 2016/679, se lleve a cabo un tratamiento con fines de investigación en salud pública y, en particular, biomédica se procederá a:

1.º **Realizar una evaluación de impacto** que determine los riesgos derivados del tratamiento en los supuestos previstos en el artículo 35 del Reglamento (UE) 2016/679 o en los establecidos por la autoridad de control. Esta evaluación incluirá de modo específico los riesgos de reidentificación vinculados a la anonimización o seudonimización de los datos.

2.º Someter la investigación científica a las **normas de calidad** y, en su caso, a las directrices internacionales sobre buena práctica clínica.

3.º Adoptar, en su caso, medidas dirigidas a **garantizar que los investigadores no acceden a datos de identificación** de los interesados.

g) El uso de datos personales seudonimizados con fines de investigación en salud pública y, en particular, biomédica deberá ser sometido al **informe previo del comité de ética de la investigación** previsto en la normativa sectorial.

Bases Jurídiques

Reial decret 957/2020, de 3 de novembre, pel qual es regulen els estudis observacionals amb medicaments d'ús humà.

Article 5. Consentiment informat i protecció de les dades personals dels subjectes participants.

1. Els estudis observacionals amb medicaments que comportin entrevistar el subjecte participant requereixen el seu consentiment informat.

No obstant això, seguint les disposicions aplicables de la normativa vigent i els principis ètics per a les recerques mèdiques en éssers humans, **es pot eximir de sol·licitar el consentiment informat, sempre que el CEIm consideri que la recerca observacional té un valor social important, que la seva realització no seria factible o viable sense aquesta dispensa, i que implica uns riscos mínims** per als participants.

Estudis observacionals amb RWD



$$\text{PRECISE NUMBER} + \text{PRECISE NUMBER} = \text{SLIGHTLY LESS PRECISE NUMBER}$$

$$\text{PRECISE NUMBER} \times \text{PRECISE NUMBER} = \text{SLIGHTLY LESS PRECISE NUMBER}$$

$$\text{PRECISE NUMBER} + \text{GARBAGE} = \text{GARBAGE}$$

$$\text{PRECISE NUMBER} \times \text{GARBAGE} = \text{GARBAGE}$$

$$\sqrt{\text{GARBAGE}} = \text{LESS BAD GARBAGE}$$

$$(\text{GARBAGE})^2 = \text{WORSE GARBAGE}$$

$$\frac{1}{N} \sum (N \text{ PIECES OF STATISTICALLY INDEPENDENT GARBAGE}) = \text{BETTER GARBAGE}$$

$$(\text{PRECISE NUMBER})^{\text{GARBAGE}} = \text{MUCH WORSE GARBAGE}$$

$$\text{GARBAGE} - \text{GARBAGE} = \text{MUCH WORSE GARBAGE}$$

$$\frac{\text{PRECISE NUMBER}}{\text{GARBAGE} - \text{GARBAGE}} = \text{MUCH WORSE GARBAGE, POSSIBLE DIVISION BY ZERO}$$

$$\text{GARBAGE} \times 0 = \text{PRECISE NUMBER}$$

Bibliografia

- **Recerca Observacional:**

- Yetley, Elizabeth & MacFarlane, Amanda & Greene-Finestone, Linda & Garza, Cutberto & Ard, Jamy & Atkinson, Stephanie & Bier, Dennis & Carriquiry, Alicia & Harlan, William & Hattis, Dale & King, Janet & Krewski, Daniel & O'Connor, Deborah & Prentice, Ross & Rodricks, Joseph & Wells, George. (2016). Options for basing Dietary Reference Intakes (DRIs) on chronic disease endpoints: report from a joint US-/Canadian-sponsored working group. American Journal of Clinical Nutrition. 105. 10.3945/ajcn.116.139097.
- Etminan M, Collins GS, Mansournia MA. Using Causal Diagrams to Improve the Design and Interpretation of Medical Research. Chest. 2020 Jul;158(1S):S21-S28. doi: 10.1016/j.chest.2020.03.011. PMID: 32658648.
- Enrique Bernal, Francisco Estupiñán. C5 Utilización de datos masivos de vida real en la investigación en servicios y políticas sanitarias, XXXIII Escola de Salut Pública de Menorca

- **Fairness of data:**

- Committee on Strategies for Responsible Sharing of Clinical Trial Data, Board on Health Sciences Policy, Institute of Medicine. (2015). COMMITTEE ON STRATEGIES FOR RESPONSIBLE SHARING OF CLINICAL TRIAL DATA
- <https://www.go-fair.org/>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). Comment: The FAIR Guiding Principles for scientific data management and stewardship Scientific Data, 3. <https://doi.org/10.1038/sdata.2016.18><https://doi.org/10.1038/sdata.2016.18>
- Javier González-Galindo, Francisco Estupiñán-Romero, David Chichell-Ruiz, Juan González-García, Carlos Tellería-Oriols, & Enrique Bernal-Delgado. (2022). PHIRI APP - WP7 - PHIRI Federated Research Infrastructure (FRI) - D7.1 Small-scale prototype of PHIRI federated infrastructure (2.2.3). Zenodo. <https://doi.org/10.5281/zenodo.6936063>

- **PGD:**

- Maseda, Irene (2020) 10 pasos para elaborar un Plan de Gestión de Datos [INFOGRAFÍA] <https://repisalud.isciii.es/handle/20.500.12105/10220>
- <https://dmptool.org/>

- **Repositoris Open Science:**

- <https://zenodo.org/>
- <https://www.openaire.eu/>
- <https://dataverse.harvard.edu/>

Bibliografia

Data Quality:

- Kotecha, D., Asselbergs, F. W., Achenbach, S., Anker, S. D., Atar, D., Baigent, C., Banerjee, A., Beger, B., Brobert, G., Casadei, B., Ceccarelli, C., Cowie, M. R., Crea, F., Cronin, M., Denaxas, S., Derix, A., Fitzsimons, D., Fredriksson, M., Gale, C. P., Grobbee, D. E. (2022). CODE-EHR best-practice framework for the use of structured electronic health-care records in clinical research. *The Lancet Digital Health* 4(10). [https://doi.org/10.1016/s2589-7500\(22\)00151-0](https://doi.org/10.1016/s2589-7500(22)00151-0)
- Sidi, Fatimah, et al. "Data quality: A survey of data quality dimensions." 2012 International Conference on Information Retrieval & Knowledge Management . 2012.
- Data Quality: The Field Guide, Thomas C. Redman, Ph.D. Digital Press, 2001
- <https://dataquality.healthdatacollaboration.net/resources#technical-resources>
- Kahn, M. G., Callahan, T. J., Barnard, J., Bauck, A. E., Brown, J., Davidson, B. N., Estiri, H., Goerg, C., Holve, E., Johnson, S. G., Liaw, S.-T., Hamilton-Lopez, M., Meeker, D., Ong, T. C., Ryan, P., Shang, N., Weiskopf, N. G., Weng, C., Zozus, M. N., & Schilling, L. (2016). A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. *EGEMs (Generating Evidence; Methods to Improve Patient Outcomes)*, 4(1), 18. <https://doi.org/10.13063/2327-9214.1244>

Imatges:

- <https://www.ibm.com/products/spss-statistics>
- <https://www.stata.com/>
- <https://www.python.org/>
- <https://pandas.pydata.org/>
- <https://shiny.rstudio.com/>
- <https://www.xkcd.com>
- <https://errantscience.com/>

Anonimitat:

- L. Sweeney. k-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10 (5), 2002; 557-570.
- Varkey, B. (2021). Principles of Clinical Ethics and Their Application to Practice. In *Medical Principles and Practice* (Vol. 30, Issue 1, pp. 17–28). S. Karger AG. <https://doi.org/10.1159/000509119>
- Thierry Meurers, Raffael Bild, Kieu-Mi Do, Fabian Prasser, A scalable software solution for anonymizing high-dimensional biomedical data, *GigaScience*, Volume 10, Issue 10, October 2021, <https://doi.org/10.1093/gigascience/giab068>

Moltes gràcies

Pau Pericàs Pulido
Responsable PRISIB
pau.pericas@ibsalut.es
idisba.prisib@ssib.es

Fundació Institut d'Investigació Sanitària Illes Balears -
IdISBa
Hospital Universitari Son Espases. Edifici " S " 1ª Planta.
Ctra. Valldemossa 79, 07120 Palma
www.idisba.es



2022 This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>.