

# Unitat de Bioestadística

Institut de Recerca Biomèdica de Lleida

## Objectius i funcionament

14 d'abril del 2020



# 1. Introducció

- La unitat de Bioestadística (UBiostat) té la missió de contribuir a generar coneixement per millorar la salut. Els mètodes estadístics són essencials per dissenyar estudis, analitzar dades i interpretar els resultats. En l'UBiostat es treballa amb investigadors per convertir les dades en informació útil, i així fer avançar la investigació. Aconsellem i animem als investigadors a consultar l'UBiostat quan es comença a planificar un estudi o projecte d'investigació.
- L'UBiostat depèn de la direcció de l'IRBLleida. Periòdicament, l'UBiostat dona comptes i presenta resultats del seu treball a la direcció de l'IRBLleida i als òrgans que la direcció considera oportuns.
- L'UBiostat té l'objectiu d'autofinançar-se amb els ingressos dels treballs realitzats i l'aportació econòmica de l'IRBLleida. Per això, s'han establert mecanismes de finançament que es descriuen en l'apartat 4 d'aquest document. Tots els grups/investigadors, tant els emergents com els consolidats que no comptin amb recursos econòmics per l'apartat estadístic, tindran garantit el reforç inicial de l'UBiostat per poder elaborar projectes d'investigació o analitzar dades que serveixin per justificar un futur finançament. Tanmateix, l'elevada demanda de serveis de l'UBiostat requereix prioritzar el reforç als grups que poden cofinançar els serveis.
- Es pot contactar amb l'UBiostat des de la pàgina web de l'IRBLleida. En l'apartat de serveis científicotècnics, dins de la pàgina de l'UBiostat, trobarà un formulari en línia per enviar la seva sol·licitud.

## 2. Objectiu i funcionament

L'**objectiu principal** de l'UBiostat és reforçar l'eficiència i qualitat de la investigació dels grups de l'IRBLleida. Aquest objectiu es concreta en les **funcions** següents:

1. Proporcionar assessorament metodològic en el disseny, anàlisi de dades i publicació d'estudis dels grups d'investigació de l'IRBLleida.
2. Participar en equips multidisciplinaris per millorar els processos de recollida, validació i integració de dades en la investigació dels grups de l'IRBLleida.
3. Participar de forma activa a les xarxes nacionals i internacionals de Bioestadística.
4. Organitzar i participar en cursos d'Estadística i metodologia de la Investigació en el context del Pla de Formació de l'IRBLleida.
5. Formar estudiants de màster i doctorat que col·laborin amb els investigadors de l'UBiostat i l'IRBLleida.
6. Proporcionar assessorament metodològic en el disseny, anàlisi de dades i publicació d'estudis a entitats externes mitjançant l'establiment de convenis de col·laboració.

## 3. Tipus de serveis

### a. Col·laboracions científiques

Participació en el disseny, anàlisi estadístic, interpretació dels resultats i redacció d'articles. Inclou la valoració del pressupost per la sol·licitud de finançament de les tasques de l'UBiostat.

### b. Reforç estadístic inicial

Assessorament en la preparació de projectes d'investigació de grups emergents i consolidats.

### c. Consultoria interna (assessorament puntual)

Permet resoldre dubtes concrets sobre algun dels aspectes de l'anàlisi estadístic d'un projecte.

### d. Formació

Consisteix en millorar la formació en Estadística dels investigadors perquè aquests siguin més autònoms. S'imparteixen cursos d'estadística bàsica, avançada i utilització del programa R per l'anàlisi de dades.

### e. Convenis

Es pretén establir col·laboracions entre l'IRBLleida i les diferents entitats per estudis i programes d'interès bilateral.

### f. Consultoria externa

L'UBiostat està oberta a demandes d'assessorament estadístic, anàlisi de dades i redacció d'informes o articles d'entitats externes.

Les sol·licituds de servei es poden fer en línia. Un cop feta la petició del servei, es concertarà una reunió amb l'usuari/a, es prepararà un pressupost i es donarà a l'usuari una data estimada de conclusió de l'estudi.

# Documents annexos

Aquest document i els annexos que s'especifiquen a continuació estan disponibles en la pàgina web de l'IRBLleida, en l'espai específic de l'UBiostat. Aquests documents s'actualitzaran periòdicament amb la finalitat d'agilitzar i millorar la comunicació entre els investigadors i l'UBiostat.

A.1 Full de sol·licitud de serveis (en línia).

A.2 Document de pla de treball, terminis i pressupost.

A.3 Aspectes ètics i confidencialitat de les dades.

A.4 Recomanacions de disseny i codificació dels arxius de dades dels estudis d'investigació.

# A.1 Full de sol·licitud de serveis

Aquest document es pot complementar en línia a la pàgina de l'UBiostat.

## Dades del sol·licitant

- Nom
- Departament/ Servei
- Institució
- Telèfon de contacte
- Correu electrònic

## Tipus d'estudi

- Amb finançament – Reforç estadístic inclòs.
- Amb finançament – Reforç estadístic no inclòs.
- Sense finançament
- Sol·licitud de nous projectes
- Altres (especificar)

## Objectiu de l'estudi

Especificar la pregunta que pretén respondre l'estudi.

## Servei/Consulta

Indicar quin tipus de consulta o reforç estadístic/ metodològic de l'UBiostat es requereix:  
(Marcar tots els apartats necessaris)

- Sol·licitud projecte competitiu
- Disseny d'estudis
- Disseny de bases de dades
- Anàlisi de dades
- Altres: especificar

## Comentaris

Text lliure

## **A.2 Document de pla de treball, terminis i pressupostos**

Aquest document l'elaboraran conjuntament el sol·licitant del servei i l'UBiostat.

### **Dades del sol·licitant**

- Sol·licitant
- Investigador principal
- Departament / Servei
- Institució
- Nom de projecte d'investigació relacionada
- Aprovat pel CEIC

### **Objectius de l'estudi**

La persona o grup que sol·licita el servei ha d'escriure amb detall els objectius de l'estudi o les preguntes que desitja respondre.

### **Pla de treball**

L'UBiostat elaborarà el Pla de Treball conjuntament amb la persona sol·licitant. El Pla de Treball ha de detallar les tasques que ses realitzaran en l'UBiostat per donar resposta a cadascun dels objectius. També hauria de detallar, si és possible, les taules i figures que hauria d'incloure l'apartat de resultats.

### **Terminis d'entrega**

L'UBiostat farà una proposta de dates d'entrega d'acord amb el volum de treball de l'UBiostat i el tipus de sol·licitud.



## **Pressupost**

L'UBiostat elaborarà un pressupost partint de la càrrega de feina que comporti la sol·licitud.

## **Firma**

Data i firma de les dues parts

## **A.3 Aspectes ètics i confidencialitat de les dades**

### **A.3.1. Aspectes ètics**

L'UBiostat tindrà en compte els aspectes ètics, la qualitat i la viabilitat de cada projecte científic, d'acord amb la legislació vigent:

- Llei 14/2007 d'investigació biomèdica.
- Llei Orgànica 15/1999 de protecció de dades de caràcter personal.
- Llei 41/2002 bàsica reguladora de l'autonomia del pacient i de drets i obligacions en matèria d'informació i documentació clínica. de bona pràctica en la investigació en ciències de la salut de l'ICS. Institut Català de la Salut, juliol de 2105

### **A.3.2. Confidencialitat de les dades**

- L'UBiostat recomana eliminar els noms dels pacients o identificadors amb el CIP o NHC dels pacients inclosos en les bases de dades que es proporcionen al personal de l'UBiostat. Tots els arxius de dades deuen incloure un codi numèric identificador del pacient.
- L'investigador principal que faciliti dades clíniques és responsable de disposar del consentiment informat dels pacients que participen en el seu projecte d'investigació.
- Tota la informació de caràcter clínic serà anònima i confidencial. L'UBiostat vetllarà per protegir la integritat i la confidencialitat de tots els arxius de dades que els investigadors li facilitin.

## **A.4 Recomanacions de disseny i codificació dels arxius de dades dels estudis d'investigació**

Aquest document pretén establir unes recomanacions que garanteixin la qualitat en el disseny de les bases de dades i la recollida de la informació. Si les dades es registren bé, l'anàlisi estadística és més ràpid i el risc de cometre errors disminueix.

### **A.4.1. Confidencialitat de les dades**

L'UBiostat recomana eliminar els noms o identificadors, com el CIP o NHC, dels pacients inclosos en les bases de dades que es proporcionen a la unitat. És necessari crear la variable identificadora (ID) que servirà per relacionar totes les taules de dades que es precisi enllaçar. El personal de l'UBiostat pot assessorar en la creació d'aquest identificador.

### **A.4.2. Disseny i gestió de bases de dades amb l'aplicació REDCap**

Per poder gestionar correctament la recollida de dades -sobretot quan el volum és important-, és recomanable utilitzar aplicacions web dissenyades per aquesta finalitat, com per exemple REDCap (Research Electronic Data Capture)

REDCap és una aplicació web segura que utilitza per la construcció i gestió d'enquestes i bases de dades en línia. REDCap pot ser utilitzada per recollir, pràcticament, qualsevol mena de dades. Està orientada, especialment, en la recollida de dades per estudis d'investigació. El consorci REDCap està format per 2.213 socis institucionals actius en 108 països, els quals utilitzen i donen reforç a REDCap de diferents formes. Actualment l'aplicació és utilitzada per més de 370.000 projectes amb més de 475.000 usuaris que abasten nombroses àrees d'interès d'investigació en tot el consorci.

Tot i requerir un cert reforç informàtic i estadístic és recomanable el seu ús per minimitzar els errors en les dades i així agilitzar l'anàlisi estadística. Per tant, us animem a contactar amb l'UBiostat abans de posar en marxa un nou estudi.

### A.4.3. Disseny i gestió de bases de dades sense aplicacions web

Si no s'utilitza cap aplicació web que garanteixi un mínim de qualitat de les dades recollides, és necessari tenir en compte certes indicacions -algunes més generals i per tant comunes a tot el conjunt de dades, i altres més específiques segons el disseny de l'estudi i la naturalesa de les dades a recollir.

- Diccionari del fitxer de dades. El diccionari del fitxer de dades serà el document que recollirà de forma exhaustiva tota la informació de les variables de l'estudi i que ajudarà a l'estadístic a analitzar les dades de forma correcta.
- Indicacions generals:
  - Cada fila de la base de dades farà referència a un registre (individu o visita) i cada columna a una variable.
  - Cada registre tindrà un identificador únic. En cas que cada individu tingui un sol registre, l'identificador es basarà en una variable (per exemple, "Pac.ID" a la Taula A.5.1). Si tenim múltiples registres pel mateix individu, l'identificador es basarà en una combinació de variables (per exemple: "Pac.ID", "Data.visita" a la Taula A.5.1).
  - No s'utilitzarà un mateix nombre per dues variables. En cas de mesurar una mateixa variable diverses vegades (per exemple: diferents visites, pre-post, ...) s'ha d'utilitzar el format de múltiples registres per un mateix individu amb l'identificador únic, així com s'indica en el punt anterior. S'haurà de registrar la data de cadascuna de les mesures.
  - Si es tracta d'una única variable mesurada dues vegades, es pot optar pel format d'un únic registre per individu. En aquest cas, les dues mesures es recullen en dues variables diferents (per exemple, "tas\_basal", "tas\_v1" a la

Figura A.5.1): la primera part del nom haurà de ser coincident entre totes i utilitzar el mateix separador (“.” o “\_”) entre el nom de la mesura (“tas”) i l’identificador del moment de la mesura (per exemple; “basal”; “v”, “post”).

- El nom de les variables no han d’incloure ni espais ni caràcters especials i hauran de reflectir el contingut de la variable de forma abreviada (per exemple: “IMC” a la Figura A.5.1). Es poden utilitzar els caràcters “.” o “\_”. (per exemple: “Pac.ID”, “Dada.visita” a la Taula A.5.1).
- Tots els valors d’una mateixa variable qualitativa hauran d’estar mesurats amb les mateixes unitats, les quals hauran d’estar identificades al diccionari de la base de dades (per exemple: “tas\_v1” en mmHg a la Figura A.5.1).
- En cas de mesures quantitatives censurades (per exemple: “Hores.d’observació” = 2, 17, 12, 5, “>24”, : : :) no es poden barrejar els valors numèrics reals amb les categories (> 24). Els valors censurats es notificaran amb codis numèrics específics sempre que sigui possible (en cap cas s’utilitzaran codis numèrics que poden ser confosos amb valors quantitius reals no censurats), o amb una variable auxiliar que indiqui les observacions censurades de la variable original.
- Totes les variables qualitatives s’hauran de recodificar numèricament i prèviament. Totes les recodificacions s’anotaran al diccionari de la base de dades (veure figura A.5.1). S’ha d’evitar l’ús d’accents i de caràcters especials (ñ,ç,º,ª,%). (per exemple: Municipi 1=Lleida, a la Taula A.5.1)
- Només en cas de preguntes obertes no recodificables (per exemple, variables de text tipus comentaris) es poden registrar com variables alfanumèriques, que en cap cas seran analitzades.
- Els valors absents (missing values) es registraran amb codis numèrics específics (s’ha d’evitar l’espai en blanc). Amb la finalitat de recollir el motiu del missing, es recomana utilitzar codis numèrics específics segons la causa del missing. Aquests codis seran registrats en el diccionari de la base de dades (veure Figura A.5.1). El codi que s’empra per indicar els missings haurà de ser

el mateix per a totes les variables de la base de dades i no podrà correspondre a cap valor possible de recodificació (per exemple: en la variable “Cig\_dia”, 998 es pot utilitzar com a valor no aplicable en cas de no fumador, mentre que 999 fer referència a valor absent (missing value))

- S’han de registrar les dates del calendari (data de naixement, data de visita, data de l’esdeveniment, ...) a partir de les quals serà possible obtenir el valor exacte de les mesures d’interès (edat d’inclusió, temps entre visites, temps de supervivència, temps fins a la recurrència,...) en les unitats temporals adequades.
- L’arxiu de la base de dades ha de gravar-se en un dels següents formats: Excel (.xls, .xlsx), SPSS (.sav) o en .csv.

Quadre A.5.1: Exemple d’estructura d’una base de dades general

Pac.ID	Data.visita	Municipi	Sexe	Edat	Fumador	Cig_dia
1	08/08/2016	1	0	41	0	998
2	10/08/2016	2	1	54	1	3

Figura A.5.1: Diccionari de la base de dades

Grup	Nom variable	Descripció variable	Tipo	Nivells/Rang	Unitats
Visita basal	Edat	Edat del pacient	Numèrica	≥ 18	
Visita basal	Sexe	Gènere del pacient	Catègòrica	0= Home 1=Dona	
Visita basal	menopausa	Si té la menopausa	Catègòrica	0=No 1=Si	
Visita basal	tas_basal	Pressió arterial sistòlica basal	Numèrica		mmHg
Visita basal	tad_basal	Pressió arterial diastòlica basal	Numèrica		mmHg
Visita basal	epworth	Índex de somnolència diürna	Numèrica	[0,24]	
Visita 1mes	IMC	Índex de massa corporal	Catègòrica	0=Bajo peso 1=Peso normal 2=Sobrepès 3=Obesitat	
Visita 1mes	hospital	Hospitalitzacions durant l’any	Catègòrica	0=No 1=Si	
Visita 1mes	hospital_obs	Observacions hospitalitzacions	Cadena		
Visita 1mes	tas_v1	Pressió arterial sistòlica 1 mes	Numèrica		mmHg
Visita 1mes	tad_v1	Pressió arterial diastòlica 1 mes	Numèrica		mmHg
...	...	...	...	...	...

#### A.4.4. Definició de les variables de la base de dades

Amb l'objectiu d'interpretar correctament la informació i facilitar l'anàlisi de les dades, és del tot necessari conèixer de forma exhaustiva cadascuna de les variables. El diccionari de la base de dades ha de ser el document que contingui la informació necessària.

- **Grup:** En cas que les variables s'estructuren en blocs, indicar a quin dels blocs pertany cada variable. (per exemple: visita basal).
- **Nom variable:** Nombrar cada una de les variables de la base de dades.
- **Descripció variable:** Descriure cadascuna de les variables de la base de dades. (per exemple, per "IMC" la descripció hauria de ser "Índex de Massa Corporal (kg/m<sup>2</sup>)")
- **Tipus:** Especificar el tipus de variable. (quantitativa, qualitativa (amb codis numèrics predefinitos sempre), o alfanumèrica sense accents ni caràcters especials).
- **Nivells/Rang:** Especificar cadascun dels nivells de la variable, en cas que sigui factor (per exemple: "nivell\_estudis", 0=sense estudis; 1=estudis primaris; 2=estudis secundaris; 3=estudis universitaris). Especificar els valors mínim i màxim possibles de la variable, en cas que sigui quantitativa (per exemple: "edat" 18, "ep-worth" 2 [0; 24]).
- **Unitats:** Especificar les unitats en la descripció de la variable, si correspon. (per exemple, per "IMC" la descripció hauria de ser "Índex de Massa Corporal (kg/m<sup>2</sup>)")