

Unitat de Bioestadística

Institut de Recerca Biomèdica de Lleida

Objectius i funcionament

31 DE MARÇ DE 2017

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. A la UBiostat es treballa amb els investigadors per convertir les dades en informació útil, i així fer avançar la recerca. Aconsellem i animem als investigadors a consultar la UBiostat quan es comença a planificar un estudi o projecte de recerca.
- La UBiostat **depèn** de la direcció de l'IRBLleida. Periòdicament, la UBiostat dona comptes i resultats de la seva tasca a la direcció de l'IRBLleida i als òrgans que la direcció consideri oportuns.
- La UBiostat **s'autofinança** amb els ingressos de les feines realitzades i el **suport** de l'IRBLleida. Es per això que s'han establert uns mecanismes de finançament que trobareu a l'apartat 4 d'aquest document. Tots els grups/investigadors, tant els emergents com els consolidats que no comptin amb recursos per a la part estadística, tindran garantit el suport inicial de la UBiostat per a poder elaborar projectes de recerca o analitzar dades que serveixin per justificar un futur finançament. No obstant això, l'elevada demanda de serveis a la UBiostat requereix prioritzar el suport als grups que poden co-finançar els serveis.
- Podeu **contactar** amb la UBiostat des de la pàgina web de l'IRBLleida. A l'apartat de serveis científic-tècnics, dins de la pàgina de la UBiostat, hi ha un formulari per enviar la sol·licitud.

2 Objectiu i funcions

L'**objectiu principal** de la UBiostat és reforçar l'eficiència i la qualitat de la recerca 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 de recerca de l'IRBLleida.
2. Participar en equips multidisciplinars per millorar els processos de recollida, validació i integració de dades en la recerca dels grups de l'IRBLleida.
3. Dur a terme recerca pròpia, tant metodològica com aplicada, en coherència amb els grups de recerca de l'IRBLleida.
4. Participar de manera activa en les xarxes nacionals i internacionals de Bioestadística.
5. Organitzar i participar en cursos d'Estadística i Metodologia de la Recerca, en el context del Pla de Formació de l'IRBLleida.
6. Formar estudiants de màster i doctorat que col·laborin amb els investigadors de la UBiostat i de l'IRBLleida.
7. 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 que s'ofereixen

1. Col·laboracions científiques

Participació en el disseny, anàlisi estadística, interpretació dels resultats i redacció d'articles. Comporta la sol·licitud de finançament per a les tasques de la UBiostat.

2. Suport estadístic inicial

Assessorament en la preparació de projectes de recerca de grups emergents i consolidats.

3. Consultoria interna (assessorament puntual)

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

4. Formació

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

5. Convenis

Es pretén establir col·laboracions entre l'IRBLleida i diferents entitats per a estudis o programes d'interés bilateral.

6. Consultoria externa

La 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. La UBiostat valorarà i prioritzarà les demandes de servei. Es prepararà un aplicatiu per tal de poder avaluar de manera automàtica el volum, procedència i tipologia de les sol·licituds de servei.

4 Co-finançament de la Unitat

Per tal de co-finançar els recursos humans de la UBioestat i fer més efectius i eficients els serveis que ofereix, s'estableix un sistema de co-pagament per part dels grups de recerca o dels investigadors individuals. S'estableixen les següents modalitats de co-finançament:

- **Grups de recerca amb necessitat continuada i intensa de suport estadístic:** finançament total o parcial de contractes de personal de la UBioestat.
- **Grups consolidats** amb necessitat de suport estadístic menys intens però continuat. Podrà establir-se una de les dues modalitats:
 - Co-finançar fraccions de contracte del personal de la unitat.
 - Co-finançar els serveis amb recursos dels projectes del grup. Si els grups no tenen finançament el personal de la UBioestat ajudarà a pressupostar les despeses en suport estadístic en els projectes futurs.
- **Grups emergents** sense projectes actius, però amb interès en desenvolupar un projecte de recerca. La UBioestat ajudarà a preparar el projecte i a sol·licitar finançament per al suport estadístic. S'oferiran 30 hores de servei gratuït i després s'aplicarà la tarifa interna que es presenta al final d'aquest apartat.
- **Convenis amb entitats externes:** la UBioestat farà una proposta basada en el volum de treball previst que es consensuarà amb la direcció de l'IRBLleida i l'entitat sol·licitant.
- **Serveis d'assessorament puntual o serveis a entitats externes sense conveni de col·laboració:** S'aplicaran tarifes variables en funció de la procedència de la demanda de servei:

	Entitat		
Dedicació	Interna ¹	Externa pública	Externa privada
1 hora	30 €	50 €	80 €

¹ Inclou IRBLleida, ICS Lleida, Gestió de Serveis Sanitaris (GSS) i UdL

5 Co-autoria del personal de la UBiostat

- Si la contribució estadística és substancial o no estàndard, la persona o persones de la UBiostat que han col·laborat en una publicació haurien de ser co-autors. Una contribució en el disseny o l'anàlisi de dades d'un estudi es considerarà substancial.
- Ser co-autor és un dret i un deure. Independentment de si es contribueix econòmicament al finançament de la UBiostat, els estadístics que col·laboren amb els grups investigadors tenen el dret i el deure, com a co-autors, d'assumir la responsabilitat que es deriva dels mètodes aplicats i la interpretació dels resultats obtinguts.
- Si el personal de la UBiostat s'inclou com co-autor en un projecte o publicació cal que doni el vist-i-plau del contingut abans d'enviar el treball. L'estadístic/a ha de tenir temps suficient per revisar i aprovar el document.

Documents annexos

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

A.1 Recursos humans de la Unitat de Bioestadística.

A.2 Full de sol·licitud de serveis (en línia).

A.3 Document de pla de treball, terminis i pressupost.

A.4 Aspectes ètics i confidencialitat de les dades.

A.5 Recomanacions de disseny i codificació dels arxius de dades dels estudis de recerca.

A.1 Recursos humans

Actualment la UBiostat està formada pels següents professionals, experts en Estadística i Metodologia de la Recerca:

- Carles Forné
- Montserrat Martínez
- Montserrat Rué
- Joan Valls

A més a més, hi ha dos estadístics adscrits al Grup de Recerca Translacional en Medicina Respiratòria:

- Iván Benítez
- Sandra Bertran

Actualment, Montserrat Rué fa les tasques següents de coordinació de la UBiostat:

- Valoració de les necessitats de serveis dels grups de recerca.
- Establiment del pla de treball i del pressupost del projecte.
- Priorització de les sol·licituds.
- Distribució de tasques entre els membres de la UBiostat.
- Seguiment dels projectes amb els grups de recerca.

A.2 Full de sol·licitud de serveis

Aquest document es pot omplir en línia a la pàgina de la UBiostat.

Dades del sol·licitant

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

Tipus d'estudi

- Amb finançament-Suport estadístic inclòs
- Amb finançament-Suport estadístic no inclòs
- Sense finançament
- Sol·licitud de nou projecte
- Altres (especificar)

Objectiu de l'estudi

Especifiqueu quina pregunta pretén respondre l'estudi

Servei/Consulta

Indiqueu quin tipus de consulta o suport estadístic/metodològic de la UBiostat necessiteu:

(Marqueu tots els apartats que calgui)

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

Comentaris

Text lliure

A.3 Document de pla de treball, terminis i pressupost

Aquest document l'elaboraran conjuntament la persona que fa la sol·licitud del servei i la UBioestat

Dades del sol·licitant

- Sol·licitant
- Investigador principal
- Departament/Servei
- Institució
- Nom del projecte de recerca relacionat
- Aprovat pel CEIC?

Objectius de l'estudi

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

Pla de Treball

La UBioestat elaborarà el pla de treball conjuntament amb la persona sol·licitant. El Pla de Treball ha de detallar les tasques que es faran a la UBioestat per donar resposta a cada-cun dels objectius. També hauria de detallar, si és possible les taules i figures que hauria d'incloure l'apartat de resultats.

Terminis d'entrega

La UBioestat farà una proposta de dates d'entrega d'acord al volum de feina de la UBioestat i el tipus de sol·licitud.

Pressupost

La UBioestat elaborarà un pressupost en base a la càrrega de treball que comporti la sol·licitud.

Signatura

Data i signatura de les dues parts

A.4 Aspectes ètics i confidencialitat de les dades

A.4.1 Aspectes ètics

La 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 dels drets i obligacions en matèria de informació i documentació clínica
- Guia de bona pràctica en la recerca en ciències de la salut de l'ICS. Institut català de la salut, juliol 2105

A.4.2 Confidencialitat de les dades

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

A.5 Recomanacions de disseny i codificació dels arxius de dades dels estudis de recerca

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 ràpida i el risc de cometre errors disminueix.

A.5.1 Confidencialitat de les dades

La UBiostat recomana eliminar els noms o identificadors, com el CIP o NHC, dels pacients inclosos en les bases de dades que es facin arribar a la unitat. Cal crear la variable identificador (ID) que servirà per relacionar totes les taules de dades que calgui enllaçar. El personal de la UBiostat pot assessorar en la creació d'aquest identificador.

A.5.2 Disseny i gestió de bases de dades amb l'aplicatiu REDCap

Per tal de gestionar correctament la recollida de dades -sobretot quan s'ha de recollir un volum important-, és recomanable fer servir aplicacions web dissenyades a tal efecte, com per exemple REDCap (Research Electronic Data Capture).

REDCap és una aplicació web segura per a la construcció i gestió d'enquestes i bases de dades en línia. REDCap pot ser utilitzada per recollir, pràcticament, qualsevol tipus de dades. Està orientada, especialment, a 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 suport a REDCap de diverses maneres. Actualment és utilitzat per més de

370.000 projectes amb més de 475.000 usuaris que engloben nombroses àrees d'interès d'investigació en tot el consorci.

Tot i que es requereix un cert suport informàtic i estadístic és recomanable el seu ús per minimitzar errors a les dades i així agilitzar les anàlisis estadístiques. Per tant, us encoratgem a contactar amb la UBiostat abans d'engegar un nou estudi.

A.5.3 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 presents certes indicacions -algunes més generals i per tant comunes a tot conjunt de dades, i d'altres més específiques segons el disseny de l'estudi i la naturalesa de les dades a recollir.

- Diccionari dels fitxers de dades. El diccionari dels fitxers 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 haurà de tenir un identificador únic. En cas que cada individu tingui un únic registre, l'identificador es basarà en una variable (per exemple, "Pac.ID" a la Taula A.5.1). Si tenim múltiples registres per a un mateix individu, l'identificador s'ha de basar en una combinació de variables (per exemple: "Pac.ID", "Data.visita" a la Taula A.5.1).
 - No s'utilitzarà un mateix nom per dues variables. En cas de mesurar una mateixa variable més d'una vegada (per exemple: diferents visites, pre-post, ...) s'ha de fer servir el format de múltiples registres per a un mateix individu amb identificador únic tal com s'indica en el punt anterior. S'ha de registrar la data de cadascuna de les mesures.

-
- Si es tracta d'una única variable mesurada dos cops, 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 ha de ser coincident entre totes i fer servir el mateix separador ("," o "_") entre el nom de la mesura ("tas") i l'identificador del moment de mesura (per exemple: "basal", "v1"; "pre", "post").
 - El nom de les variables no ha d'incloure espais ni caràcters especials i haurà de reflectir el contingut de la variable de forma abreujada (per exemple: "IMC" a la Figura A.5.1). Es poden fer servir els caràcters "." o "_". (per exemple: "Pac.ID", "Data.visita" a la Taula A.5.1)
 - Tots els valors d'una mateixa variable quantitativa hauran d'estar mesurats amb les mateixes unitats, les quals hauran d'indicar-se 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.observacio" = 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 puguin ser confosos amb els 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 es faran constar 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, °, 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 comentari) es poden deixar com a variables alfanumèriques, que en cap cas seran analitzades.
 - Els valors absents (*missings values*) es notificaran amb codis numèrics específics (s'ha d'evitar l'espai en blanc). Per tal de recollir el motiu del *missing*, es recomana fer servir codis numèrics específics segons la causa del *missing*. Aquests codis seran registrats al diccionari de la base de dades (veure Figura

A.5.1). El codi que s'utilitzi 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: a la variable "Cig_dia", 998 es pot utilitzar com a valor no aplicable en cas dels no fumadors, mentre que 999 fa referència a valor absent (*missing value*))

- S'han de registrar les dates del calendari (data de naixement, dates de les visites, dates dels esdeveniments, ...) a partir de les quals serà possible obtenir els valors exactes de les mesures d'interès (edat d'inclusió, temps entre visites, temps de supervivència, temps fins a la recurrència, ...) en les unitats temporals adients.
- L'arxiu de la base de dades s'ha de gravar en un dels següents formats: Excel (.xls, .xlsx), SPSS (.sav) o en .csv.

Taula 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	Tipus	Nivells / Rang				Unitats
Visita basal	edat	Edat del pacient	Numèrica	≥ 18				
Visita basal	sexe	Gènere del pacient	Catègorica	0=Home	1=Dona			
Visita basal	menopausia	Si té la menopausia	Catègorica	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	Catègorica	0 = Pes sota del normal	1 = Pes normal	2 = Sobrepès	3 = Obesitat	
Visita 1mes	hospital	Hospitalitzacions l'últim any	Catègorica	0 = No	1 = Si			
Visita 1mes	hospital_obs	Observacions hospitalització	Cadena					
Visita 1mes	tas_v1	Pressió arterial sistòlica 1 mes	Numèrica					mmHg
Visita 1mes	tad_v1	Presión arterial diastòlica 1 mes	Numèrica					mmHg
...	...	...	...	...	...	...	...	...

A.5.4 Definició de les variables de la base de dades

Per tal 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 tota la informació necessària.

-
- **Grup:** En cas que les variables s'estructurin en blocs, indicar a quin bloc pertany cada variable. (per exemple: visita basal)
 - **Nom variable:** Anomenar cadascuna de les variables de la base de dades.
 - **Descripció variable:** Descriure cadascuna de les variables de la base de dades. (per exemple, per a "IMC" la descripció hauria de ser "Índex de Massa Corporal (kg/m2)")
 - **Tipus:** Especificar el tipus de variable. (quantitativa, qualitativa (amb codis numèrics predefinits 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 el mínim i màxim possible de la variable, en cas que sigui quantitativa (per exemple: "edat" ≥ 18 , "epworth" $\in [0, 24]$).
 - **Unitats:** Especificar les unitats a la descripció de la variable, si s'escau. (per exemple, per a "IMC" la descripció hauria de ser "Índex de Massa Corporal (kg/m2)")