

# Unidad de Bioestadística

Instituto de Investigación Biomédica de  
Lleida

## *Objetivos y funcionamiento*

31 DE MARZO DE 2017

# 1 Introducción

- La Unidad de Bioestadística (UBiostat) tiene la **misión** de contribuir a generar conocimiento para mejorar la salud. Los métodos estadísticos son esenciales para diseñar estudios, analizar datos e interpretar los resultados. En la UBiostat se trabaja con los investigadores para convertir los datos en información útil, y así hacer avanzar la investigación. Aconsejamos y animamos a los investigadores a consultar la UBiostat cuando se empieza a planificar un estudio o proyecto de investigación.
- La UBiostat **depende** de la dirección del IRBLleida. Periódicamente, la UBiostat da cuentas y presenta resultados de su labor a la dirección del IRBLleida y a los órganos que la dirección considera oportunos.
- La UBiostat **se autofinancia** con los ingresos de los trabajos realizados y la **aportación económica** del IRBLleida. Por ello, se han establecido mecanismos de financiación que se describen en el apartado 4 de este documento. Todos los grupos/investigadores, tanto los emergentes como los consolidados que no cuenten con recursos económicos para el apartado estadístico, tendrán garantizado el apoyo inicial de la UBiostat para poder elaborar proyectos de investigación o analizar datos que sirvan para justificar una futura financiación. Sin embargo, la elevada demanda de servicios a la UBiostat requiere priorizar el apoyo a los grupos que pueden cofinanciar los servicios.
- Se puede **contactar** con la UBiostat desde la página web del IRBLleida. En el apartado de servicios científico-técnicos, dentro de la página de la UBiostat, encontrará un formulario en línea para enviar su solicitud. se encuentra un formulario para enviar la solicitud.

## 2 Objetivo y funcionamiento

El **objetivo principal** de la UBiostat es reforzar la eficiencia y la calidad de la investigación de los grupos del IRBLleida. Este objetivo se concreta en las **funciones** siguientes:

1. Proporcionar asesoramiento metodológico en el diseño, análisis de datos y publicación de estudios de los grupos de investigación del IRBLleida.
2. Participar en equipos multidisciplinares para mejorar los procesos de recolección, validación e integración de datos en la investigación de los grupos del IRBLleida.
3. Realizar investigación propia, tanto metodológica como aplicada, en coherencia con los grupos de investigación del IRBLleida.
4. Participar de forma activa en las redes nacionales e internacionales de Bioestadística.
5. Organizar y participar en cursos de Estadística y Metodología de la Investigación, en el contexto del Plan de Formación del IRBLleida.
6. Formar estudiantes de máster y doctorado que colaboren con los investigadores de la UBiostat y del IRBLleida.
7. Proporcionar asesoramiento metodológico en el diseño, análisis de datos y publicación de estudios a entidades externas mediante el establecimiento de convenios de colaboración.

## 3 Tipo de servicios

### 1. Colaboraciones científicas

Participación en el diseño, análisis estadístico, interpretación de los resultados y redacción de artículos. Incluye la valoración del presupuesto para solicitud de financiación de las tareas de la UBiostat.

### 2. Apoyo estadístico inicial

Asesoramiento en la preparación de proyectos de investigación de grupos emergentes y consolidados.

### 3. Consultoría interna (asesoramiento puntual)

Permite resolver dudas concretas sobre alguno de los aspectos del análisis estadístico de un proyecto.

### 4. Formación

Consiste en mejorar la formación en Estadística de los investigadores para que éstos sean más autónomos. Se imparten cursos de Estadística básica, avanzada y utilización del programa R para el análisis de datos.

### 5. Convenios

Se pretende establecer colaboraciones entre el IRBLleida y diferentes entidades para estudios o programas de interés bilateral.

### 6. Consultoría externa

La UBiostat está abierta a demandas de asesoramiento estadístico, análisis de datos y redacción de informes o artículos de entidades externas.

Las solicitudes de servicio se pueden hacer en línea. La UBiostat valorará y priorizará las demandas de servicio. Se preparará un aplicativo para poder evaluar de manera automática el volumen, procedencia y tipología de las solicitudes de servicio.

## 4 Cofinanciación de la Unidad

Para ofertar un servicio más efectivo y eficiente, se establece un sistema de cofinanciación de los recursos humanos de la UBiostat. Se establecen las siguientes modalidades de cofinanciación:

- **Grupos de investigación con necesidad continuada e intensa de apoyo estadístico:** financiación total o parcial de contratos de personal de la UBiostat.
- **Grupos consolidados** con necesidad de apoyo estadístico menos intenso pero continuado. Podrá establecerse una de las dos modalidades:
  - Cofinanciar fracciones de contrato del personal de la unidad.
  - Cofinanciar los servicios con recursos de los proyectos del grupo. Si los grupos no tienen financiación el personal de la UBiostat ayudará a presupuestar los gastos en apoyo estadístico en los proyectos futuros.
- **Grupos emergentes** sin proyectos activos, pero con interés en desarrollar un proyecto de investigación. La UBiostat ayudará a preparar el proyecto y a solicitar financiación para el apoyo estadístico. Se ofrecerán 30 horas de servicio gratuito y después se aplicará la tarifa interna que se presenta al final de este apartado.
- **Convenios con entidades externas:** La UBiostat hará una propuesta basada en el volumen de trabajo previsto que se consensuará con la dirección del IRBLleida y la entidad solicitante.
- **Servicios de asesoramiento puntual o servicios a entidades externas sin convenio de colaboración:** Se aplicarán tarifas variables en función de la procedencia de la demanda de servicio:

	Entidad		
Dedicación	Interna <sup>1</sup>	Externa pública	Externa privada
1 hora	30 €	50 €	80 €

<sup>1</sup>Incluye IRBLleida, ICS Lleida, Gestió de Serveis Sanitaris (GSS) y UdL

## 5 Coautoría del personal de la UBiostat

- Si la contribución estadística es sustancial o no estándar, la persona o personas de la UBiostat que han colaborado en una publicación deberían ser coautores. Una contribución en el diseño o el análisis de datos de un estudio se considerará sustancial.
- Ser coautor es un derecho y un deber. Independientemente de si se contribuye económicamente a la financiación de la UBiostat, los estadísticos que colaboran con los grupos investigadores tienen el derecho y el deber, como coautores, de asumir la responsabilidad que se deriva de los métodos aplicados y la interpretación de los resultados obtenidos.
- Si el personal de la UBiostat se incluye como coautor en un proyecto o publicación es necesario que esté de acuerdo con el contenido antes de ser enviado el trabajo. El estadístico/a debe tener tiempo suficiente para revisar y aprobar el documento.

# Documentos anexos

Este documento y los anexos que se especifican a continuación están disponibles en la página web del IRBLleida, en el espacio específico de la UBiostat. Estos documentos se actualizarán periódicamente con el fin agilizar y mejorar la comunicación entre los investigadores y la UBiostat.

A.1 Recursos humanos de la Unidad de Bioestadística.

A.2 Hoja de solicitud de servicios (en línea).

A.3 Documento de plan de trabajo, plazos y presupuesto.

A.4 Aspectos éticos y confidencialidad de los datos.

A.5 Recomendaciones de diseño y codificación de los archivos de datos de los estudios de investigación.

# A.1 Recursos humanos

Actualmente la UBiostat está formada por los siguientes profesionales, expertos en Estadística y Metodología de la Investigación:

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

Además, hay dos estadísticos adscritos al Grupo de Investigación Traslacional en Medicina Respiratoria:

- Iván Benítez
- Sandra Bertran

Actualmente, Montserrat Rué hace las tareas siguientes de coordinación de la UBiostat:

- Valoración de las necesidades de servicios de los grupos de investigación.
- Establecimiento del plan de trabajo y del presupuesto del proyecto.
- Priorización de las solicitudes.
- Distribución de tareas entre los miembros de la UBiostat.
- Seguimiento de los proyectos con los grupos de investigación.



## A.2 Hoja de solicitud de servicios

Este documento se puede cumplimentar en línea en la página de la UBiostat.

### Datos del solicitante

- Nombre
- Departamento/Servicio
- Institución
- Teléfono de contacto
- Correo electrónico

### Tipo de estudio

- Con financiación -Apoyo estadístico incluido
- Con financiación -Apoyo estadístico no incluido
- Sin financiación
- Solicitud de nuevos proyecto
- Otros (especificar)

### Objetivo del estudio

Especificar la pregunta que pretende responder el estudio.

---

## Servicio/Consulta

Indicar qué tipo de consulta o apoyo estadístico/metodológico de la UBiostat se requiere:

(Marcar todos los apartados necesarios)

- Solicitud proyecto competitivo
- Diseño de estudios
- Redacción de artículos
- Diseño de bases de datos
- Análisis de datos
- Otros: especificar

## Comentarios

Texto libre

## **A.3 Documento de plan de trabajo, plazos y presupuesto**

Este documento lo elaborarán conjuntamente el/la solicitante del servicio y la UBiostat.

### **Datos del solicitante**

- Solicitante
- Investigador principal
- Departamento/Servicio
- Institución
- Nombre del proyecto de investigación relacionado
- Aprobado por el CEIC?

### **Objetivos del estudio**

La persona o grupo que solicita el servicio debe describir con detalle los objetivos del estudio o las preguntas que desea responder.

### **Plan de Trabajo**

La UBiostat elaborará el Plan de Trabajo conjuntamente con la persona solicitante. El Plan de Trabajo debe detallar las tareas que se realizarán en la UBiostat para dar respuesta a cada uno de los objetivos. También debería detallar, si es posible, las tablas y figuras que debería incluir el apartado de resultados.

---

## **Plazos de entrega**

La UBiostat hará una propuesta de fechas de entrega de acuerdo al volumen de trabajo de la UBiostat y el tipo de solicitud.

## **Presupuesto**

La UBiostat elaborará un presupuesto en base a la carga de trabajo que conlleve la solicitud.

## **Firma**

Fecha y firma de las dos partes

## A.4 Aspectos éticos y confidencialidad de los datos

### A.4.1. Aspectos éticos

La UBioestat tendrá en cuenta los aspectos éticos, la calidad y la viabilidad de cada proyecto científico, de acuerdo con la legislación vigente:

- Ley 14/2007 de investigación biomédica
- Ley Orgánica 15/1999 de protección de datos de carácter personal
- Ley 41/2002 básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica
- Guía de buena práctica en la investigación en ciencias de la salud del ICS. Institut Català de la Salut, julio de 2105

### A.4.2. Confidencialidad de los datos

- La UBioestat recomienda eliminar los nombres de los pacientes o identificadores como el CIP o NHC de los pacientes incluidos en las bases de datos que se proporcionen al personal de la UBioestat. Todos los archivos de datos deben incluir un código numérico identificador del paciente.
- El investigador principal que facilite datos clínicos es responsable de disponer del consentimiento informado de los pacientes que participan en su proyecto de investigación.
- Toda la información de carácter clínico será anónima y confidencial. La UBioestat velará para proteger la integridad y la confidencialidad de todos los archivos de datos que los investigadores le faciliten.

## **A.5 Recomendaciones de diseño y codificación de los archivos de datos de los estudios de investigación**

Este documento pretende establecer unas recomendaciones que garanticen la calidad en el diseño de las bases de datos y la recogida de la información. Si los datos se registran bien, el análisis estadístico es más rápido y el riesgo de cometer errores disminuye.

### **A.5.1. Confidencialidad de los datos**

La UBiostat recomienda eliminar los nombres o identificadores, como el CIP o NHC, de los pacientes incluidos en las bases de datos que se proporcionen a la unidad. Es necesario crear la variable identificador (ID) que servirá para relacionar todas las tablas de datos que se precise enlazar. El personal de la UBiostat puede asesorar en la creación de este identificador.

### **A.5.2. Diseño y gestión de bases de datos con la aplicación REDCap**

A fin de gestionar correctamente la recogida de datos -sobre todo cuando el volumen es importante-, es recomendable utilizar aplicaciones web diseñadas para tal fin, como por ejemplo REDCap (Research Electronic Data Capture).

REDCap es una aplicación web segura que se utiliza para la construcción y gestión de encuestas y bases de datos en línea. REDCap puede ser utilizada para recoger, prácticamente, cualquier tipo de datos. Está orientada, especialmente, a la recogida de datos

---

para estudios de investigación. El Consorcio REDCap está formado por 2.213 socios institucionales activos en 108 países, los cuales utilizan y dan apoyo a REDCap de distintas maneras. Actualmente la aplicación es utilizada por más de 370.000 proyectos con más de 475.000 usuarios que abarcan numerosas áreas de interés de investigación en todo el consorcio.

A pesar de requerir un cierto apoyo informático y estadístico es recomendable su uso para minimizar los errores en los datos y así agilizar el análisis estadístico. Por lo tanto, os animamos a contactar con la UBiostat antes de poner en marcha un nuevo estudio.

### **A.5.3. Diseño y gestión de bases de datos sin aplicaciones web**

Si no se utiliza ninguna aplicación web que garantice un mínimo de calidad de los datos recogidos, es necesario tener en cuenta ciertas indicaciones -algunas más generales y por lo tanto comunes a todo conjunto de datos, y otras más específicas según el diseño del estudio y la naturaleza de los datos a recoger.

- Diccionario del fichero de datos. El diccionario del fichero de datos será el documento que recogerá de forma exhaustiva toda la información de las variables del estudio y que ayudará al estadístico a analizar los datos de forma correcta.
- Indicaciones generales:
  - Cada fila de la base de datos hará referencia a un registro (individuo o visita) y cada columna a una variable.
  - Cada registro tendrá un identificador único. En caso de que cada individuo tenga un solo registro, el identificador se basará en una variable (por ejemplo, "Pac.ID" en la Tabla A.5.1). Si tenemos múltiples registros para el mismo individuo, el identificador se basará en una combinación de variables (por ejemplo: "Pac.ID", "Data.visita" a la Tabla A.5.1).
  - No se utilizará un mismo nombre para dos variables. En caso de medir una misma variable distintas veces (por ejemplo: diferentes visitas, pre-post, ...) se

---

debe utilizar el formato de múltiples registros para un mismo individuo con el identificador único tal y como se indica en el punto anterior. Se deberá registrar la fecha de cada una de las medidas.

- Si se trata de una única variable medida dos veces, se puede optar por el formato de un único registro por individuo. En este caso, las dos medidas se recogen en dos variables distintas (por ejemplo: “tas\_basal”, “tas\_v1” a la Figura A.5.1): la primera parte del nombre tendrá que ser coincidente entre todas y utilizar el mismo separador (“.” o “\_”) entre el nombre de la medida (“tas”) y el identificador del momento de la medida (por ejemplo: “basal”, “v1”; “pre”, “post”).
- El nombre de las variables no deben incluir espacios ni caracteres especiales y deberá reflejar el contenido de la variable de forma abreviada (por ejemplo: “IMC” a la Figura A.5.1). Se pueden utilizar los caracteres “.” o “\_”. (por ejemplo: “Pac.ID”, “Data.visita” a la Tabla A.5.1)
- Todos los valores de una misma variable cualitativa deberán estar medidos con las mismas unidades, las cuales deberán estar identificadas en el diccionario de la base de datos (por ejemplo: “tas\_v1” en mmHg a la Figura A.5.1).
- En caso de medidas cuantitativas censuradas (por ejemplo: “Horas.observacion” = 2, 17, 12, 5, “>24”, ...) no se pueden mezclar los valores numéricos reales con las categorías (> 24). Los valores censurados se notificarán con códigos numéricos específicos siempre que sea posible (en ningún caso se usarán códigos numéricos que puedan ser confundidos con valores cuantitativos reales no censurados), o con una variable auxiliar que indique las observaciones censuradas de la variable original.
- Todas las variables cualitativas se deberán recodificar numéricamente y previamente. Todas las recodificaciones se anotarán en el diccionario de la base de datos (ver Figura A.5.1). Se debe evitar el uso de acentos y de caracteres especiales (ñ, ç, °, ª, %). (por ejemplo: Municipio 1=Lleida, a la Tabla A.5.1)
- Solamente en caso de preguntas abiertas no recodificables (por ejemplo, variables de texto tipo comentarios) se pueden registrar como variables alfanuméricas, que en ningún caso serán analizadas.



- Los valores ausentes (*missings values*) se registrarán con códigos numéricos específicos (se debe evitar el espacio en blanco). Con el fin de recoger el motivo del *missing*, se recomienda utilizar códigos numéricos específicos según la causa del *missing*. Estos códigos serán registrados en el diccionario de la base de datos (ver Figura A.5.1). El código que se emplee para indicar los *missings* deberá ser el mismo para todas las variables de la base de datos y no podrá corresponder a ningún valor posible de recodificación. (por ejemplo: en la variable “Cig\_dia”, 998 se puede utilizar como valor no aplicable en caso de no fumador, mientras que 999 hace referencia a valor ausente (*missing value*))
- Se deben registrar las fechas del calendario (fecha de nacimiento, fecha de visita, fecha del evento, ...) a partir de las cuales será posible obtener el valor exacto de las medidas de interés (edad de inclusión, tiempo entre visitas, tiempo de supervivencia, tiempo hasta la recurrencia, ...) en las unidades temporales adecuadas.
- El archivo de la base de datos debe grabarse en uno de los siguientes formatos: Excel (.xls, .xlsx), SPSS (.sav) o en .csv.

Cuadro A.5.1: Ejemplo de estructura de una base de datos general

Pac.ID	Data.visita	Municipio	Sexo	Edad	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: Diccionario de la base de datos

Grupo	Nombre variable	Descripción variable	Tipo	Niveles/Rango	Unidades
Visita basal	edad	Edad del paciente	Numérica	≥ 18	
Visita basal	sexo	Género del paciente	Catógica	0=Hombre 1=Mujer	
Visita basal	menopausia	Si tiene la menopausia	Catógica	0=No 1=Si	
Visita basal	tas_basal	Presión arterial sistólica basal	Numérica		mmHg
Visita basal	tad_basal	Presión arterial diastólica basal	Numérica		mmHg
Visita basal	epworth	Índice de somnolencia diurna	Numérica	[0,24]	
Visita 1mes	IMC	Índice de masa corporal	Catógica	0=Bajo peso 1=Peso normal 2=Sobrepeso 3=Obesidad	
Visita 1mes	hospital	Hospitalizaciones durante el año	Catógica	0=No 1=Si	
Visita 1mes	hospital_obs	Observaciones hospitalizaciones	Cadena		
Visita 1mes	tas_v1	Presión arterial sistólica 1 mes	Numérica		mmHg
Visita 1mes	tad_v1	Presión arterial diastólica 1 mes	Numérica		mmHg
...	...	...	...	...	...

---

## A.5.4. Definición de las variables de la base de datos

A fin de interpretar correctamente la información y facilitar el análisis de los datos, es del todo necesario conocer de forma exhaustiva cada una de las variables. El diccionario de la base de datos debe ser el documento que contenga la información necesaria.

- **Grupo:** En caso de que las variables se estructuren en bloques, indicar a cuál de los bloques pertenece cada variable. (por ejemplo: visita basal)
- **Nombre variable:** Nombrar cada una de las variables de la base de datos.
- **Descripción variable:** Describir cada una de las variables de la base de datos. (por ejemplo, para “IMC” la descripción debería ser “Índice de Masa Corporal (kg/m<sup>2</sup>)”)
- **Tipo:** Especificar el tipo de variable. (cuantitativa, cualitativa (con códigos numéricos predefinidos siempre), o alfanumérica sin acentos ni caracteres especiales)
- **Niveles/Rango:** Especificar cada uno de los niveles de la variable, en caso que sea factor (por ejemplo: “nivel\_estudios”, 0=sin estudios; 1=estudios primarios; 2=estudios secundarios; 3=estudios universitarios). Especificar los valores mínimo y máximo posibles de la variable, en caso que sea cuantitativa (por ejemplo: “edad”  $\geq 18$ , “ep-worth”  $\in [0, 24]$ ).
- **Unidades:** Especificar las unidades en la descripción de la variable, si corresponde. (por ejemplo, para “IMC” la descripción debería ser “Índice de Masa Corporal (kg/m<sup>2</sup>)”)