

Professional life of a young biostatistician

Susana Pérez-Álvarez

Societat Catalana d'Estadística
Institut de Recerca de la SIDA – IrsiCaixa

Jornades de Consultoria Estadística i Software II
23 d'Octubre de 2013

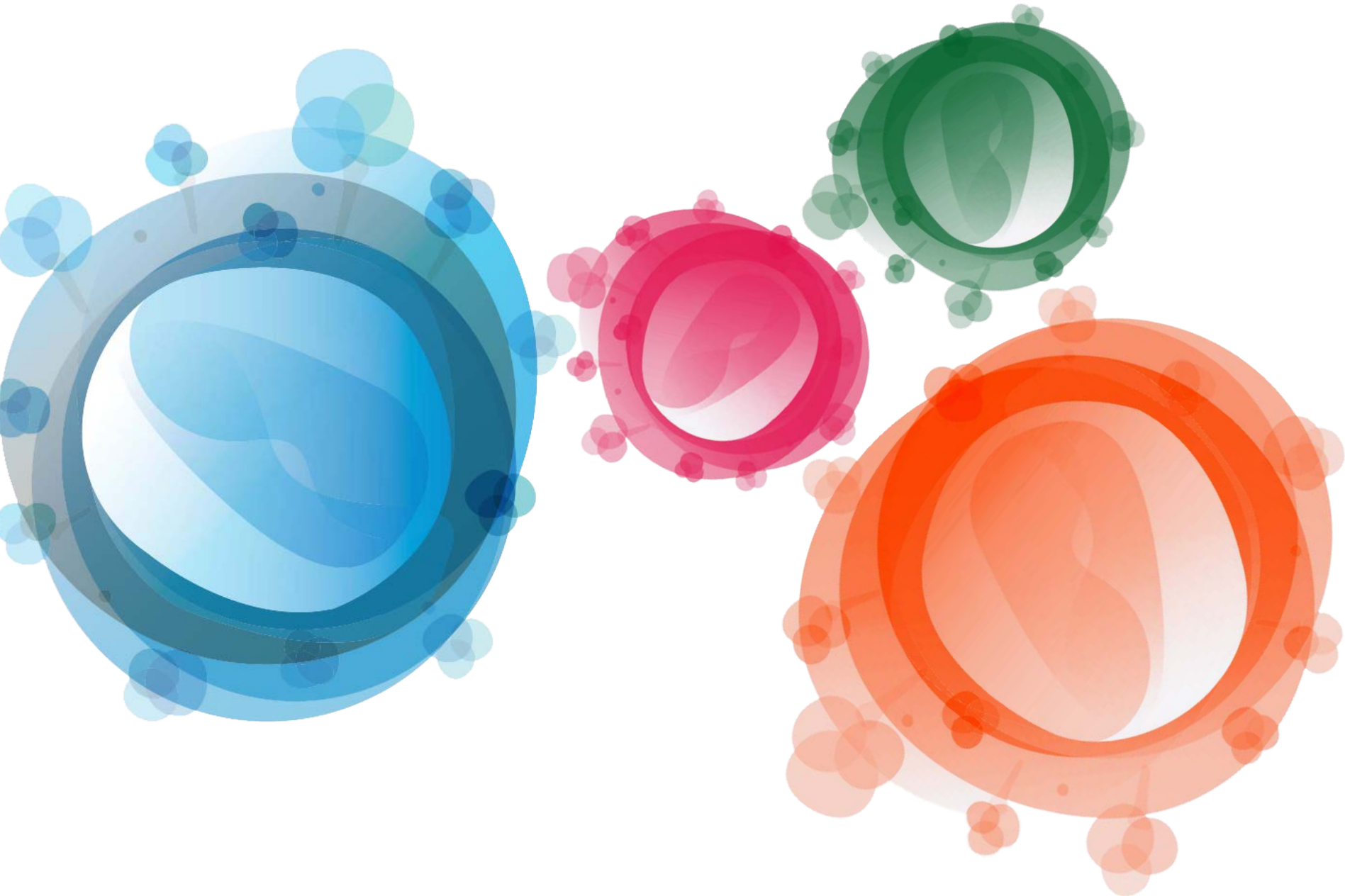


Outline

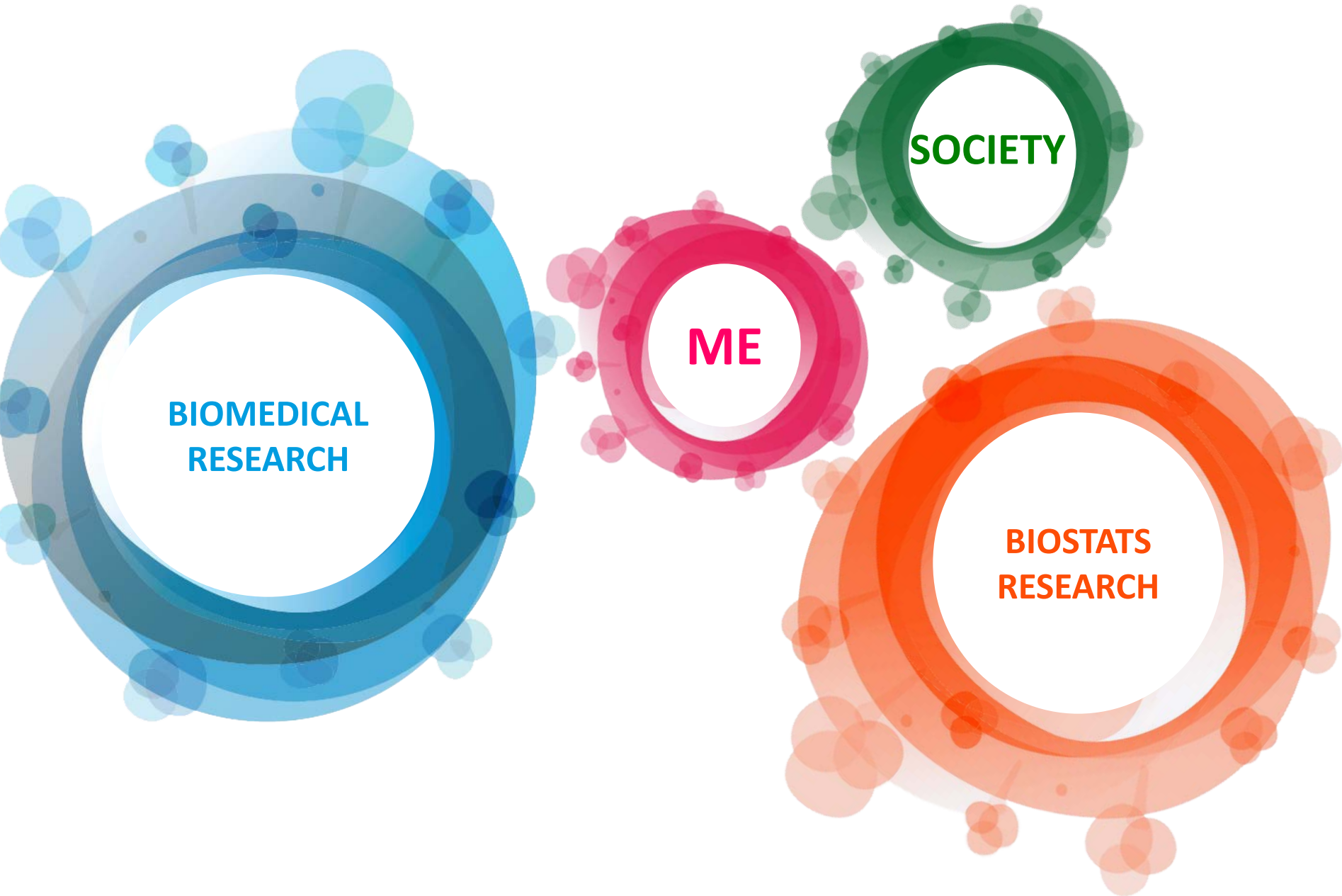
- Me, myself and life
- Once upon a time...society
- An then, there was research
- Biostats happily ever after!



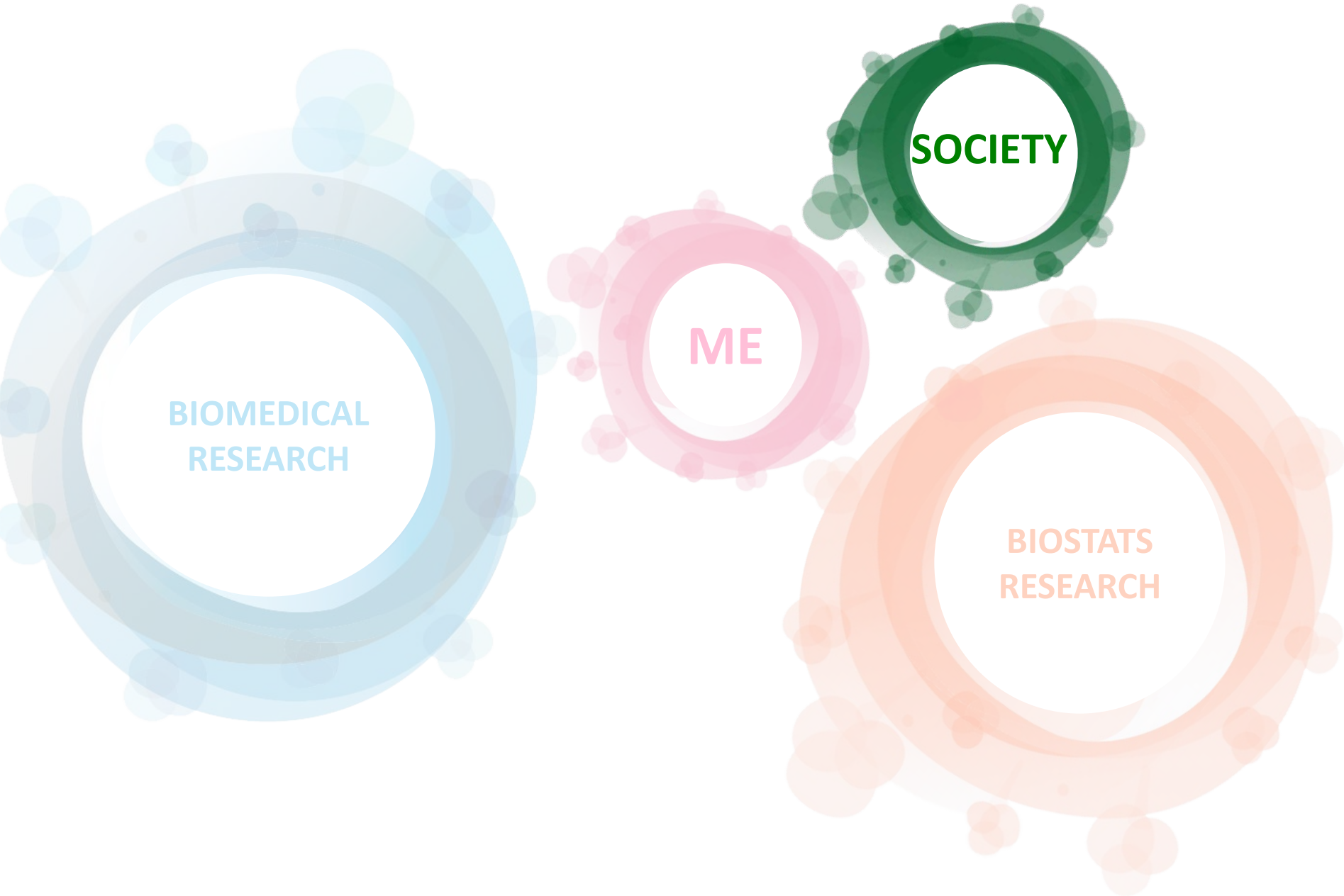
Me, myself and life



Me, myself and life



Once upon a time...Society



Once upon a time...Society



Societat Catalana
d'Estadística – SCE

www.socestadistica.org

Once upon a time...Society



Societat Catalana
d'Estadística – SCE

www.socestadistica.org

- Founded in February 2010
- 180 members
- Secretary and board member since October 2012
- Board members:
 - Guadalupe Gómez – UPC
 - Xavier López – Idescat
 - Pere Puig – UAB
 - Joan Guàrdia – UB
 - Rosa Lamarca – Almirall
 - Ernest Pons – UB
 - Xavier Tort-Martorell – UPC

Once upon a time...Society



Societat Catalana
d'Estadística – SCE

www.socestadistica.org

- Founded in February 2010
- 180 members
- Secretary and board member since October 2012
- Management tasks
 - Minutes and Assemblies
 - Membership
 - Members communication
- Activities organization
 - Statistical Analysis Contest
- Monthly bulletin
- “Networking”
- Be “in society”

Once upon a time...Society

SOCIETAT CATALANA D'ESTADÍSTICA



Bossa d'Ajuts per a cursos d'estadística impartits a Catalunya

2013

Què et cal per rebre l'ajut?

- Ser membre de la SCE
- Presentar el programa, durada, entitat organitzadora i preu del curs a la secretaria de la SCE:
secretari@socestadistica.org
- Es concedirà un màxim d'un ajut per membre.

Termini per sol·licitar l'ajut: 15 de juny de 2013

Resolució ajut: 30 de juny de 2013

MÉS INFORMACIÓ A: WWW.SOCESTADISTICA.ORG

Grant Eligibility :

- The student must be a member of the SCE
- Maximum 1 grant per member
- Submit the program, duration, organizer and course fee information in the secretariat of the SCE
- Courses can not be part of the official courses offered in bachelor or master degree, except those that are part of the summer programs.

Once upon a time...Society

Concurso de análisis estadístico 2013

Convoca
Sociedad Catalana
de Estadística

Colabora
The Data Republic

Calendario:

- 15 Mayo 2013
Inicio de las inscripciones
- 30 Octubre 2013
Finaliza el plazo de entrega de trabajos
- Noviembre 2013
Selección de los trabajos finalistas
- Diciembre 2013
Jornada de presentación de los trabajos finalistas y entrega de premio al ganador

Más información en:
www.socestadistica.org

SOE THE DATA REPUBLIC

Statistical Analysis Contest:

- Dead-line: 31st October 2013
- Data from “The Data Republic”
- Winner award: Ipad.
- Jury:
 - 1 member SCE
 - 1 representative TDR
 - 1 statistician
- Resolution: 11th December 2013

Once upon a time...Society

e-NEWS
SocE
Núm. 10 - Octubre de 2013

Hola Mireia, si vols aportar algun comentari o suggeriment respecte a l'e-NEWS fes clic aquí

NOTÍCIES SCE

Jornada de celebració de l'Any Internacional al Parlament
El proper divendres 18 d'Octubre es celebrarà l'Any Internacional de l'Estadística amb un acte al Parlament de Catalunya convidats per la seva presidenta, la M. H. Sra. Núria de Gispert. La jornada començarà a les 9.30h del matí i s'allargarà fins al migdia.
[Més informació, en un clic.](#)

Dia de l'Estadística a Catalunya
El passat 4 d'Octubre vam celebrar el dia de l'estadística a Catalunya a l'Institut d'Estudis Catalans. Hi ha haver una gran afluència de públic que va omplir la Sala. La jornada va començar recordant altres celebracions del dia de l'estadística, va introduir-se al món bayesià de la mà d'en David Rossell i va seguir amb la taula rodona titulada "Que pot fer l'estadística per la salut?". A més, el públic va participar al trivial estadístic i l'equip guanyador es va emportar un gran premi!
[Més informació, en un clic.](#)

Assemblea Ordinària
A continuació de la celebració del dia de l'estadística, va tenir lloc l'Assemblea Ordinària. Entre altres, un punt del dia va ser l'elecció dels càrrecs de dos vocals. En Daniel Baena i en Vladimír Zalts deixen la Junta de la SCE i prenen el seu relleu en Joan Guàrdia i en Xavier Tort-Martorell. Moltes gràcies al Dani i al Vladimír per tota la feina feta, i benvinguts al Joan i al Xavier!

"L'Alfabetització estadística. Quin paper hi juguem els ensenyants?"
Sota aquest títol es va celebrar la 10a Jornada d'Ensenyament de les Matemàtiques el passat 28 de Setembre a l'IEC. A la taula rodona que es va fer al matí, hi va participar la presidenta de la SCE així com altres membres de la nostra societat. Al link següent hi podeu trobar el programa i el material de la taula rodona i les ponències.
[Més informació, en un clic.](#)

ACTIVITATS de la SCE

Concurs d'anàlisi estadística - 2013
El passat 15 de Maig es va obrir el període d'inscripcions per al concurs d'anàlisi estadística que hem convocat amb la col·laboració de *The Data Republic*. El premi és un iPad i el termini per presentar els treballs acaba el 30 d'Octubre. Animeu-vos a participar així com a fer-ne la màxima difusió possible!
[Més informació, en un clic.](#)

Bloc e2013
Aquest mes d'Octubre el bloc porta el títol "L'estadística edificant", on experts de l'IEWAE (Institut d'Estadística i Matemàtica aplicada a l'Edificació), entre altres professionals, ens parlaran de l'estadística al camp de l'edificació.
[Feu una visita al bloc i animeu-vos a fer un comentari!](#)

Seminari "Survival times using mixture models (competing risks methods)"

[Per a més informació un clic](#)

biologia de la
a el patró de
amb Almirall,
licada a l'Any
l'Indústria
ou brut de
arc Científic i
sió al voltant
presentants i
lari, mèdic,
del Dr. Xavier
ial Statistics.
da de la UAB
ió d'aquestes
nt de trobada
ir-se en una
ientífica com
OE, però els
d'anglès i
te laboral
tat de fer
Pontificia
the world,
or about
of people
as the top
oinformàtica
nternacional,
periòdiques,
de la unitat i
periència en
rporar al teu
consisteix en
Es valoraran
de bases de
eficis socials.

e-NEWS SocE

- Monthly bulletin
- Sections such as:
 - Society News
 - Statistical News
 - Activities organized by SCE
 - Other activities
 - Jobs
 - Conferences
 - Contests and Awards

Once upon a time...Society

- **Jornada de celebració de l'Any internacional de l'estadística**

Friday 18th



Once upon a time...Society



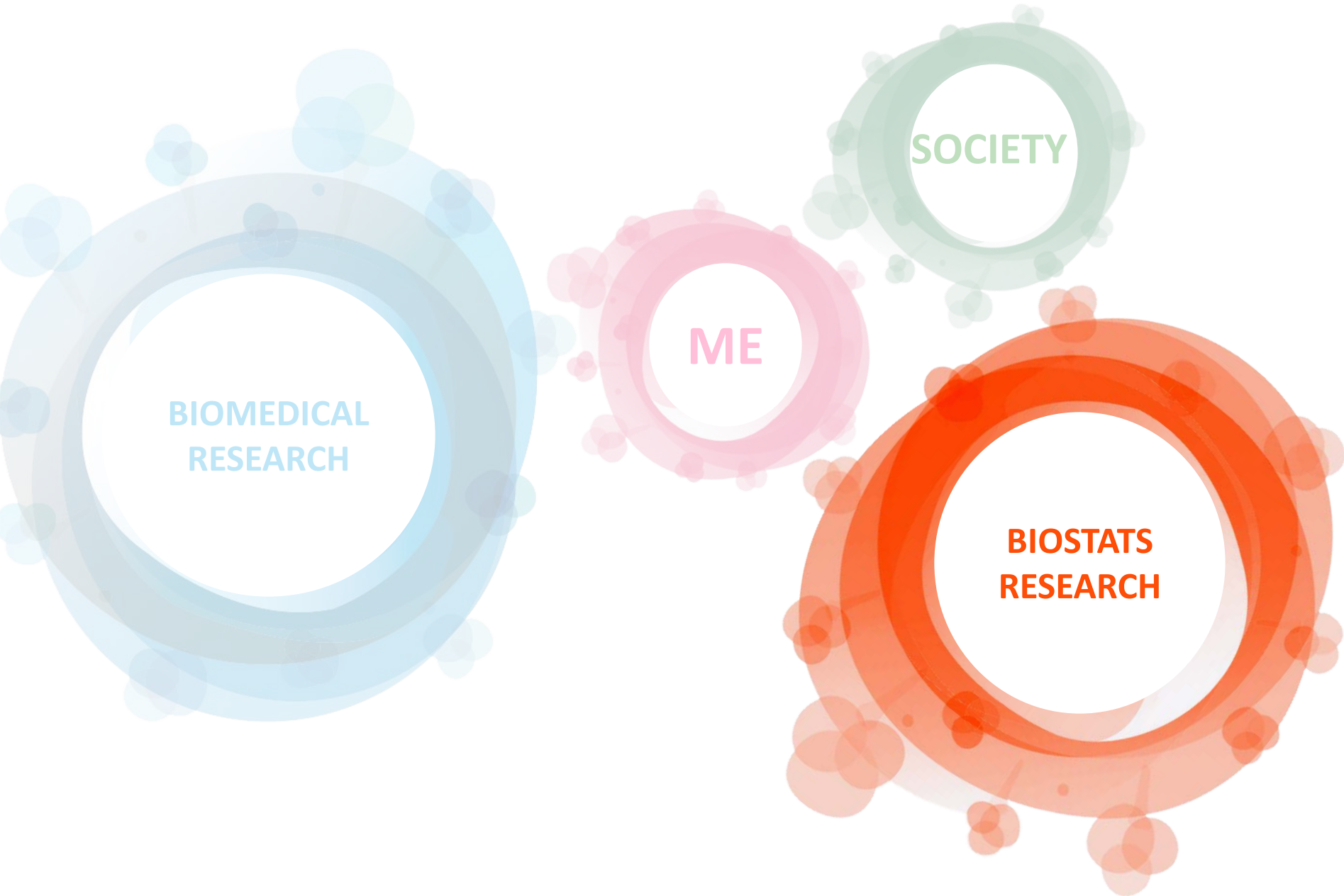
Societat Catalana
d'Estadística – SCE

www.socestadistica.org

JOIN US!

- Regular fee: 50€
- Reduced fees:
 - Undergraduate and graduate students: 20€
 - Under 30 year: 20€
 - Pensioners and retirees: free
 - Unemployed over thirty years: 20€
- Sponsors: 500€ or higher

An then, there was research



An then, there was research



Grup de **R**ecerca en **A**nàlisi
estadística de la **S**upervivència

An then, there was research



- Member since 2005
 - BSc thesis development
- Objectives
 - To foment and spread research into the **Survival Analysis**.
 - To **help** investigators and applied research groups to use this kind of techniques.
 - To design and offer **tools and documentation** to these groups in order to help its practice application in several fields.
 - To establish **a meeting point** to investigators who are interested in these techniques.

An then, there was research



- Members:

- Guadalupe Gómez Melis
- M. Luz Calle Rosingana
- Carles Serrat Piè
- Montse Rué Monné
- Núria Porta Bleda
- Ramon Oller Piqué
- Klaus Langohr
- Olga Julià de Ferran
- Jaime Abel Huertas Campos
- Anna Espinal Berenguer

- Víctor Urrea Gales
- Albert Roso Llorach
- Oleguer Plana Ripoll
- Nuria Pérez Álvarez
- Lucas López Segovia
- Raquel López Blazquez
- Moisés Gómez-Mateu
- Carles Forné Izquierdo
- Carolina Barroeta Rojo

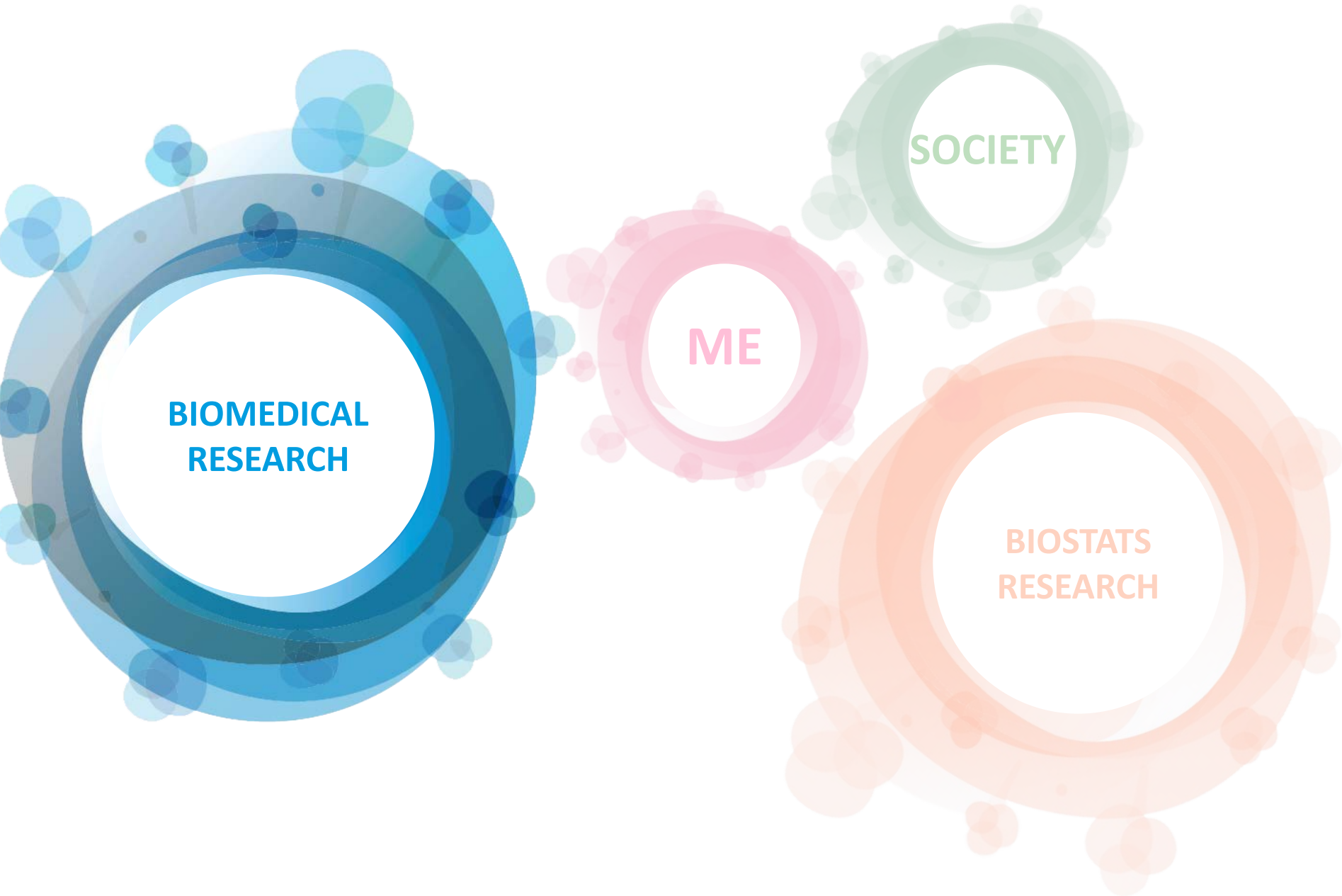
An then, there was research



Thematic network aiming to link together Spanish researchers in biostatistics with an integrative, multidisciplinary, flexible, and open focus.

- Pioneer network in Spain
 - 8 nodes led by statisticians from different universities
- Gathers and represents the main lines of investigation of each of the nodes involved.
- Objectives are:
 - a) Coordinate research and teaching of biostatistics;
 - b) Promote an adequate education in biostatistics;
 - c) Help to show its transparency and applicability in biomedicine.

Biostats happily ever after!



Biostats happily ever after!

IrsiCaixa

Institut de Recerca de la Sida

HIVACAT

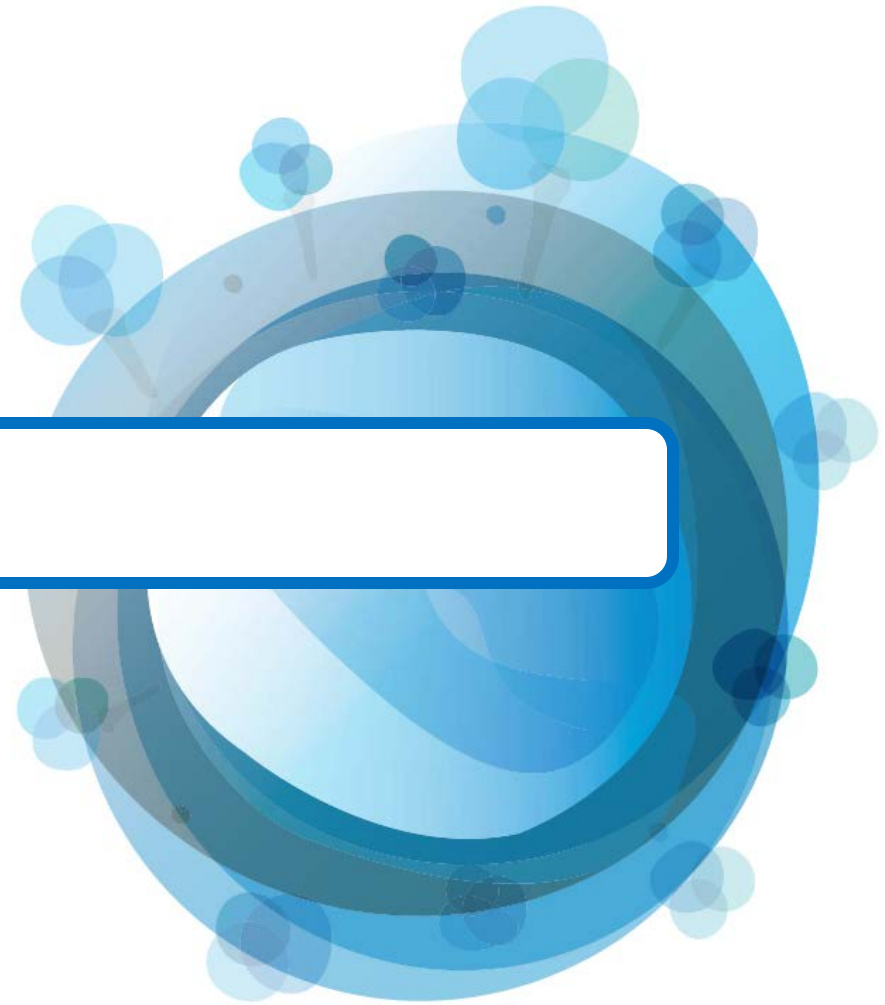
Projecte de Recerca de la Vacuna de la Sida

- Host Cellular and Genetic Immunity
 - Dr. Christian Brander

- Molecular Epidemiology
 - Dr. Roger Paredes

Project 01

FARMS



FARMS

- **F**orward and
- **A**ll-Subsets
- **R**egression for
- **M**odel
- **S**election
- Based on...

Liebminger et. al. "A new variable selection method based on All Subsets Regression " 10th Scandinavian Symposium on Chemometrics, 2007.

FARMS main features

- Composing the starting model
 - Forced in covariates
 - Covariates that will always be in the model
 - Randomly select covariates
- Select the best subset of covariates
 - R^2 , Adjusted R^2 , Mallows' C_p , AIC or BIC
- Select the best model
 - AIC or BIC

FARMS data and parameters

- Data
 - Y response variable
 - P covariables
- Parameters
 - P_a number of covariables to add on a step
 - P_M maximum number of covariables in the model
- Optional parameters
 - P_f covariables forced to be in the final model
 - P_s covariables in the starting model

FARMS applied to HIV data

- Viral Load of 867 HIV infected patients
 - 236 patients from Lima, Peru, South America
 - 631 patients from Durban, South Africa
- Immune Data
 - 73 variables corresponding to HLA subtypes
 - 406 variables corresponding to OLP

FARMS evaluation

- Evaluate all possible combinations when...
 - $P_{\text{start}} = 2, 3, 4, \dots, 20, 21$
 - $P_{\text{add}} = 2, 3, 4, \dots, 20, 21$
- Always obtained the same model
 - HLAs covariates:
 - 8 covariates
 - $R^2=11.67\%$
 - $\text{Adj.}R^2=10.74\%$
 - OLPs covariates
 - 11 covariates
 - $R^2=14.77\%$
 - $\text{Adj.}R^2=13.57\%$
- Total time and number of iterations varied

FARMS applied on HIV

		FARMS	All-Subsets	Forward Selection	Forward Stepwise	Backward Stepwise
HLA	Time* <i>Mean (IQR)</i>	1.26s (1.01; 1.40)	>1week	1.99s (1.61; 1.88)	2.38s (1.92; 2.22)	13s
	Num. Of Covariates**	9	-	10	9	10
	BIC	2235.44	-	2235.45	2235.44	2236.52
	AIC	2183.03	-	2178.27	2183.03	2174.58
	R ²	11.67%	-	12.36%	11.67%	12.93%
	Adj. R ²	10.74%	-	11.33%	10.74%	11.81%
OLP	Time* <i>Mean (IQR)</i>	28.48s (19.17; 30.80)	>1week	38.02s (31.34; 33.08)	38.25 (31.34; 33.38)	>12hours
	Num. Of Covariates**	12	-	12	12	23
	BIC	2224.82	-	2224.82	2224.82	2232.63
	AIC	2158.11	-	2158.11	2158.11	2108.63
	R ²	14.77%	-	14.77%	14.77%	21.68%
	Adj. R ²	13.57%	-	13.57%	13.57%	19.45%

* Including "forced in" covariates

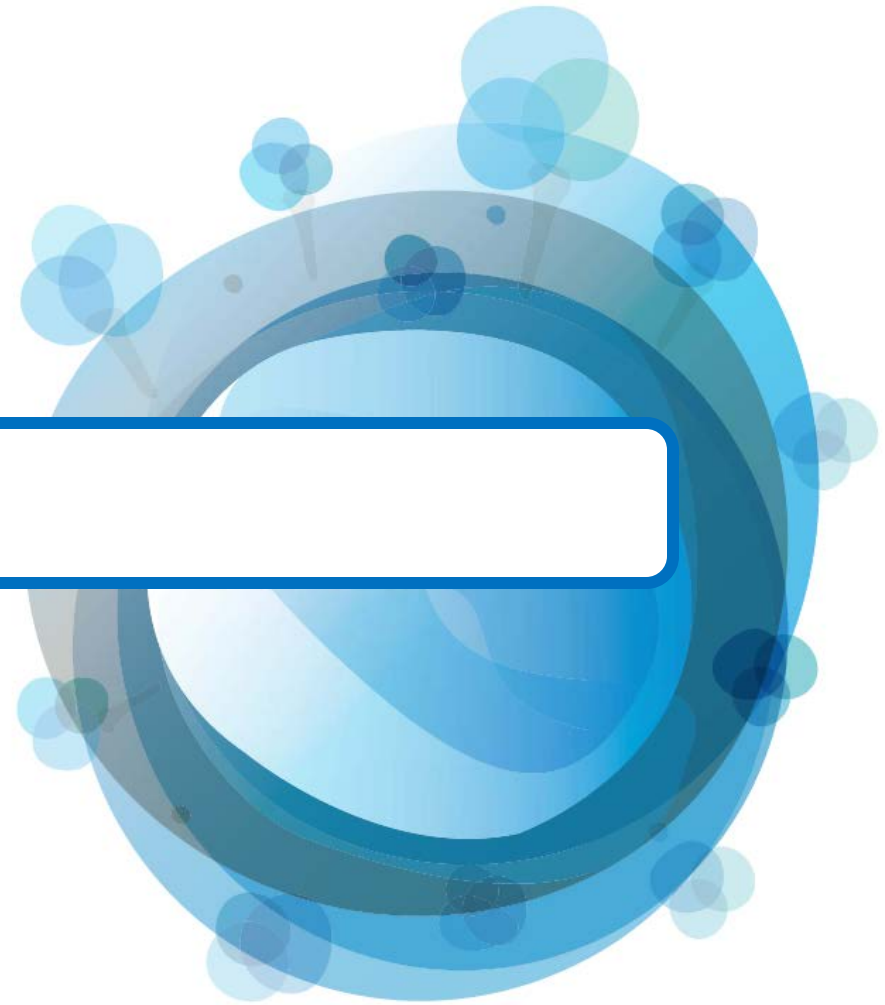
** Pa=10; Ps=8; Selecting criterion = BIC

FARMS Conclusions

- FARMS allows more control on selecting a model according to datasets
- FARMS is faster than other common approaches to model/variable selection
- We obtained more parsimonious models
- Future work:
 - Compare with other methods: LASSO, LARS, ...
 - Improve subset selection within FARMS inner-loop

Project 02

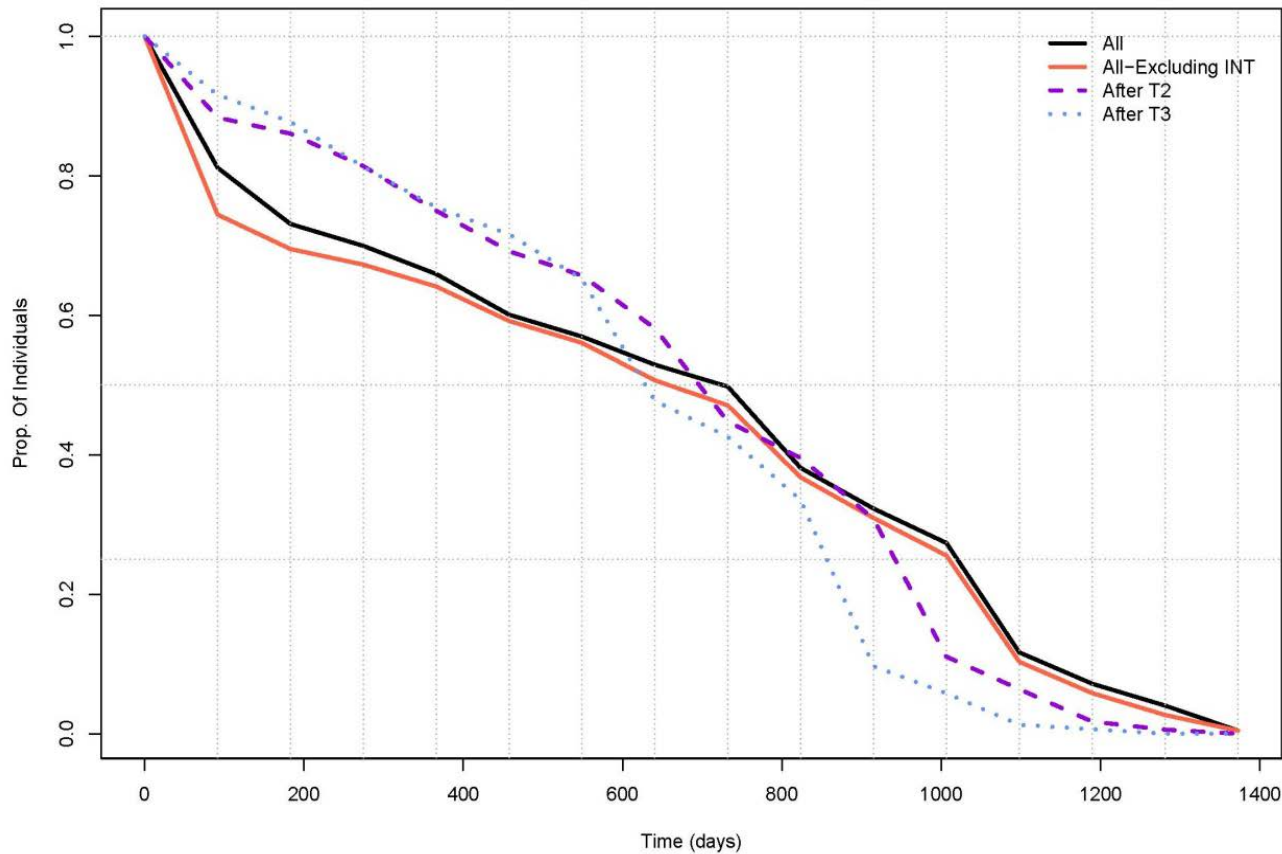
LIMA COHORT



Lima cohort (I): risk factors

- 243 subjects enrolled in Lima (Peru)
 - High risk cohort
 - MSM: men who have sex with men
 - Extensive questionnaires every 3 months
- We look for factors influencing:
 - Recruitment
 - Retention
 - Seroconversion

Lima cohort (I): Adherence

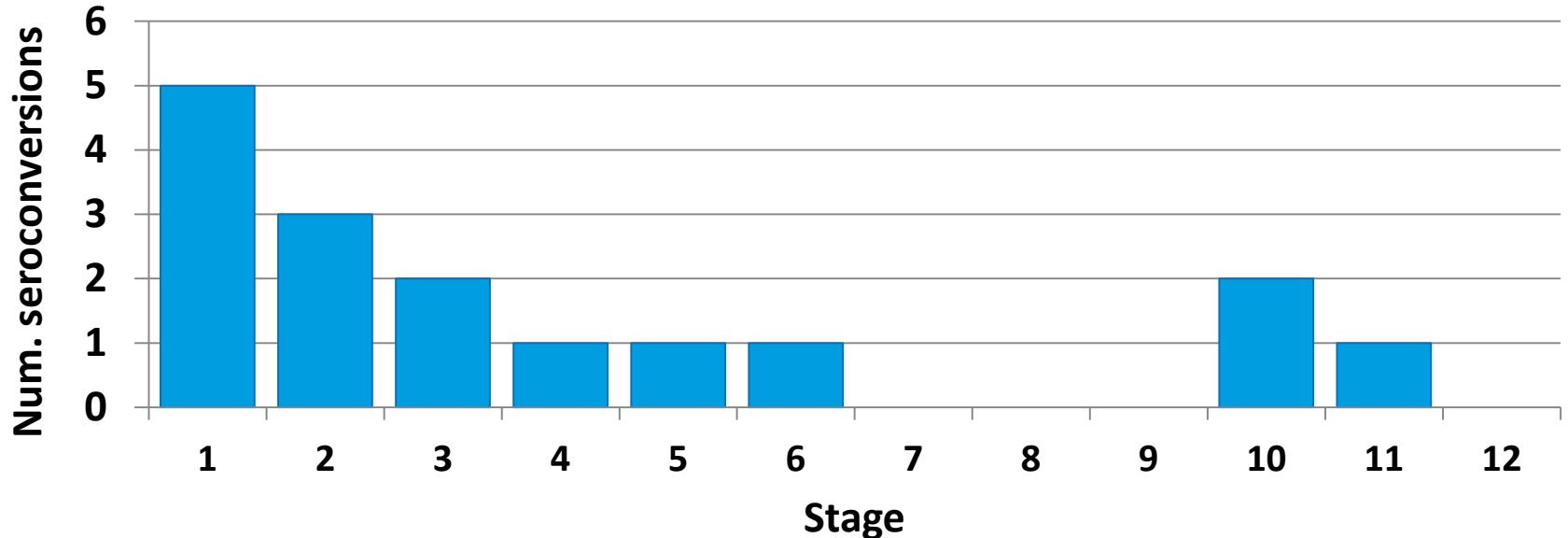


Proportion of people included in the cohort

Each line represents inclusion of all visits to clinic, only study visits (excluding intermediate visits), after visit 1 (excluding patients only have 1 visit) or after visit 2 (excluding patients only have 2 visits).

- The sharp drop-off in follow-up between 1st and 2nd/3rd visit and the relative lack of new infections at later time points may help to **optimize cohort management**

Lima cohort (I): Serconversion rate



- Lack of late seroconversions may reflect an **educational effect** among the enrollees during their participation in these studies.

Lima cohort (I): Conclusions

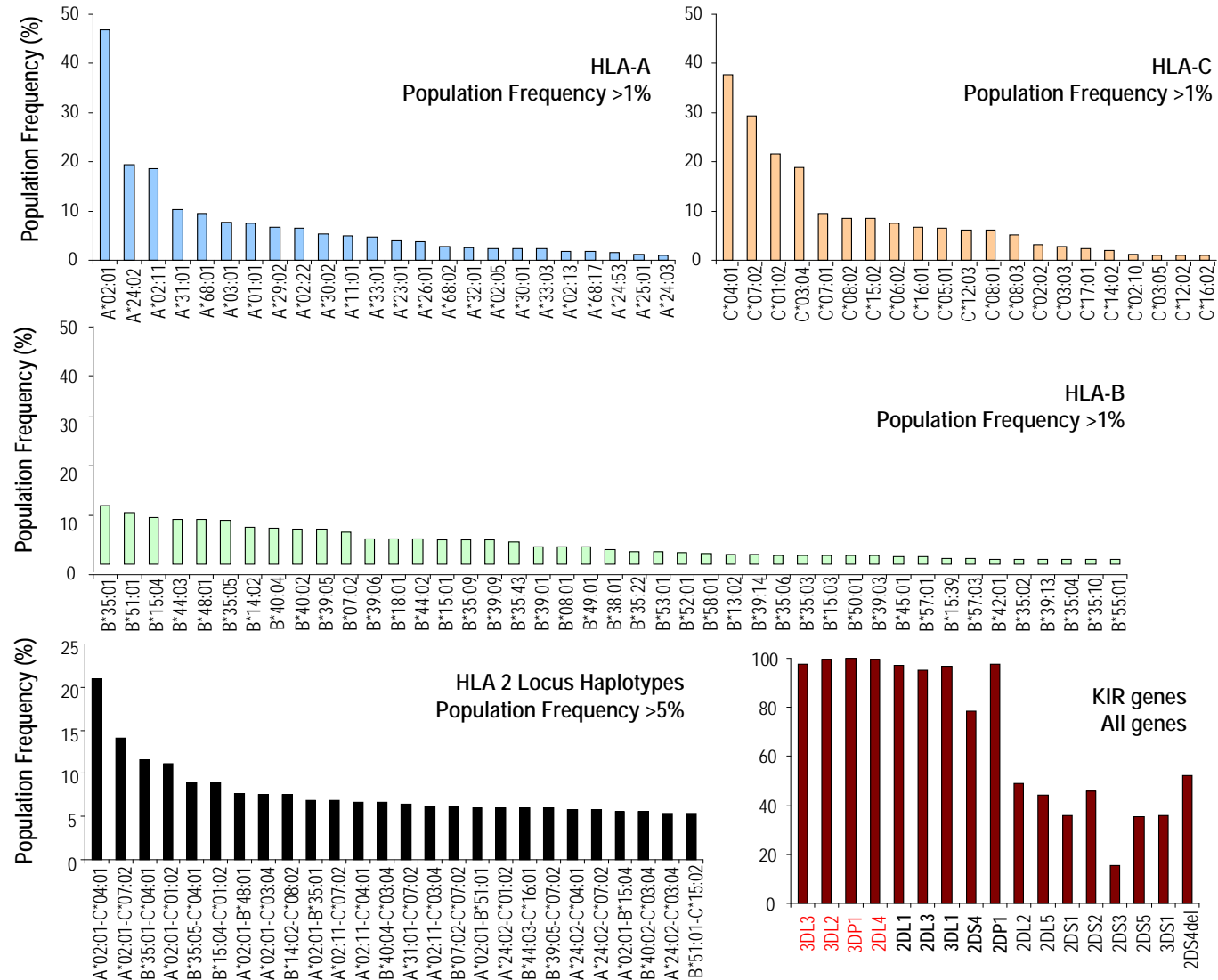
- The establishment of high risk cohorts is **feasible but poses high demands**
 - on enrollees, study staff and lab infrastructure
- Advertising, enrolment and retention strategies have to be **tailored to the specific target population** and to **involve the community**
- Close follow up of individuals at high risk of HIV infection allows us to **have a group of candidates prepared to participate in HIV vaccine clinical trials**
- The sharp drop-off in follow-up may help to **optimize cohort management**
 - **shorter duration**
 - **focussed follow-up**
- Lack of late seroconversions may also reflect an **educational effect** among the enrollees during their participation in these studies.

Lima cohort (II): Disease progression

- Assess potential associations between Human Leukocyte Antigens (HLA) and Killer-cell Immunoglobulin-like Receptors (KIR) and the risk of HIV infection, HIV viral load or CD4 counts
- **468 individuals** with high risk sexual behavior recruited and tested in Lima, Peru
 - 222 HIV negative
 - 246 HIV positive
- Each subject was **4-digit typed for HLA** class I alleles
- **KIR genotypes**
 - Available for 73 HIV– and 170 HIV+ subjects
- **Viral loads** and **CD4 counts** in infected individuals

Lima cohort (II): Disease progression

Cohort Description

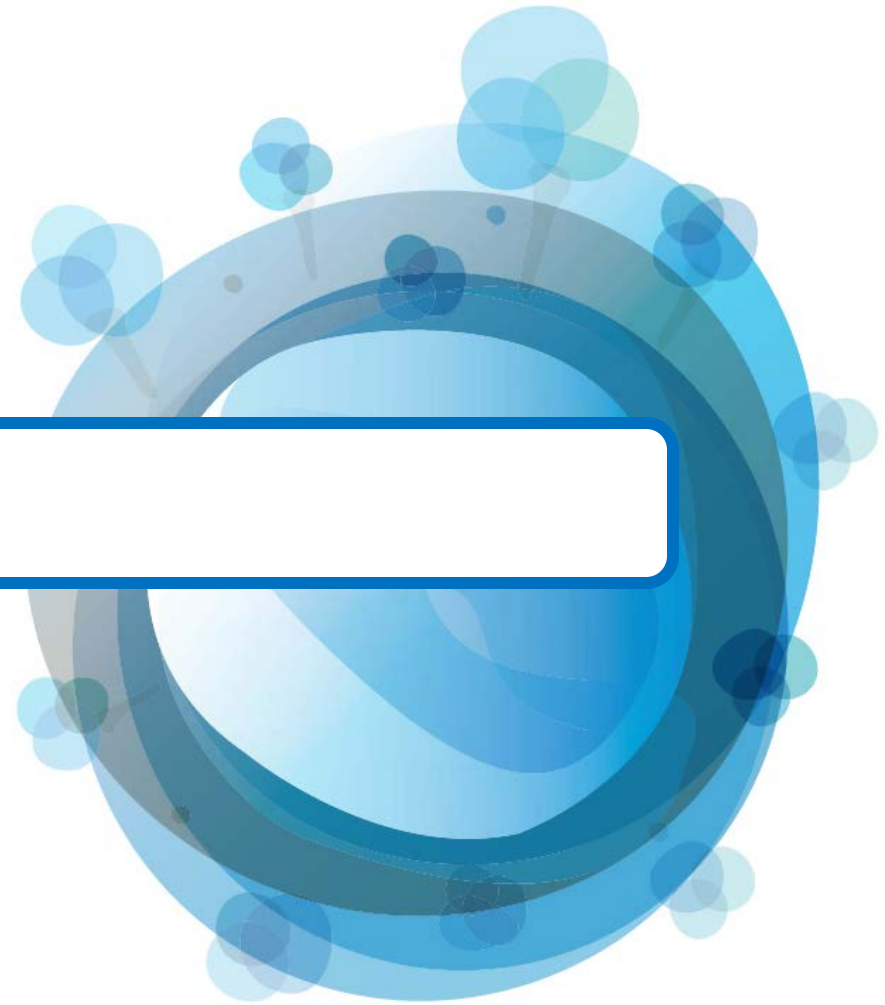


Lima cohort (II): Conclusions

- Despite Peruvian cohorts being well represented in past and current HIV vaccine trials, this is the **first study to define high-resolution** HLA and KIR typing for this population.
- The description of the genetic background of this population will be of importance for immunopathological studies in this population and to interpret disease outcome and HIV vaccine trial data in this country and the rest of South America.

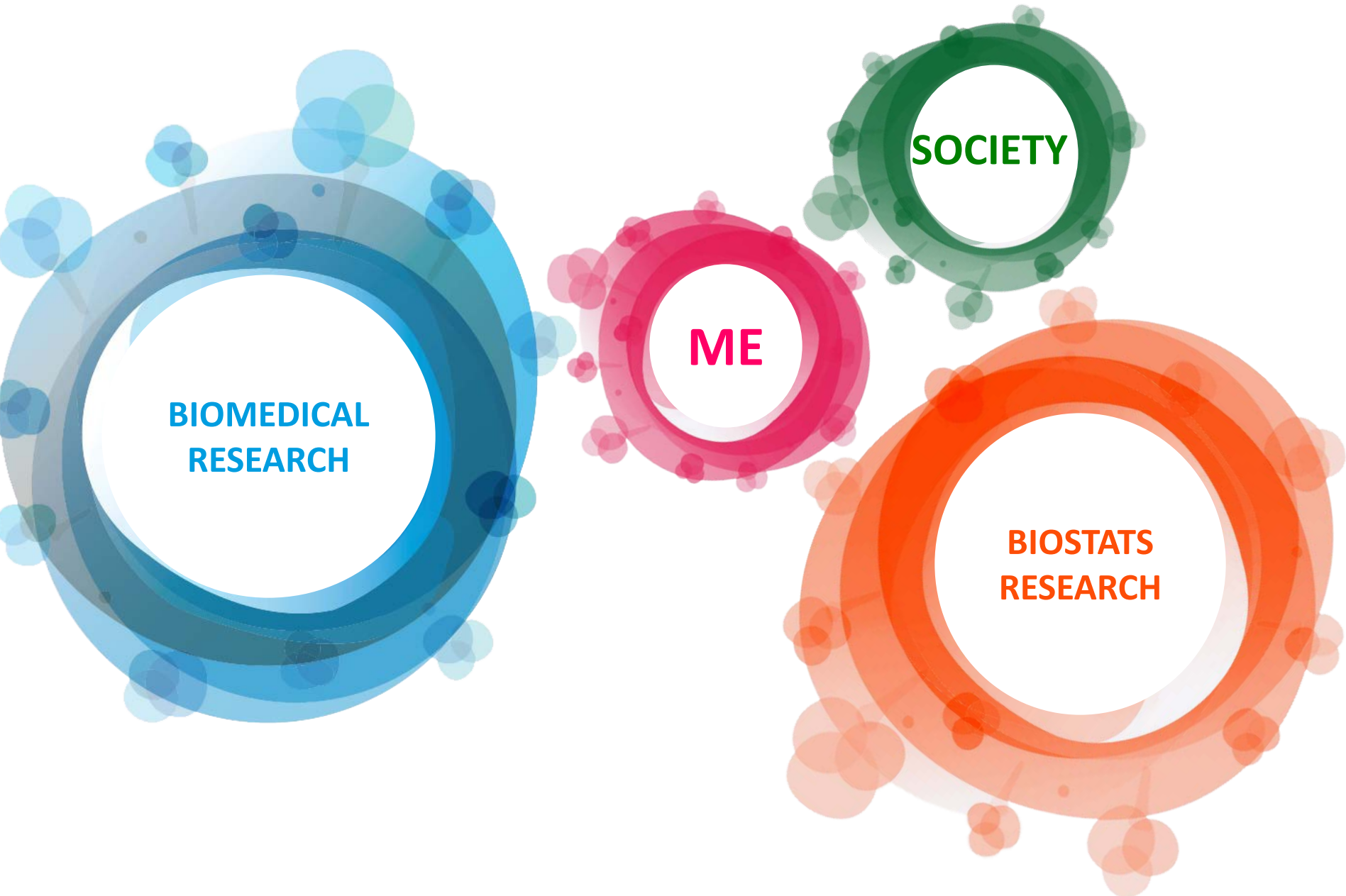
Project 03

PRIUS STUDY



PRIUS Study Overview

- **GOAL:** Evaluate the impact of minority variants in drug-resistance and virological failure
 - Compare two different sequencing technologies
- Design: Observational, multicenter, retrospective cohort study.
- Participant centers:
 - Hospital 12 de Octubre, Madrid
 - Mútua de Terrassa, Terrassa
 - Hospital Universitario San Cecilio, Granada
 - Laboratorio de Retrovirología irsiCaixa, Badalona



Acknowledgements



Institut de Recerca de la Sida

HIVACAT

Projecte de Recerca de la Vacuna de la Sida



**AÑO INTERNACIONAL
DE LA ESTADÍSTICA**



THANKS

FOR

YOUR

ATTENTION

Professional life of a young biostatistician

Susana Pérez-Álvarez

Societat Catalana d'Estadística
Institut de Recerca de la SIDA – IrsiCaixa

Jornades de Consultoria Estadística i Software II
23 d'Octubre de 2013

