EURORDIS
Rare Diseases Europe

EURORDIS MEMBERSHIP MEETING
DUBROVNIK 2013
30May-1June, Hotel Rixos Libertas

Data and cues for a European Platform for rare diseases registration

Fabrizio Bianchi

Unit of Epidemiology, Institute of Clinical Physiology, CNR
Tuscany Registry of Rare Diseases at G. Monasterio Foundation
Epirare WP6 leader

and EPIRARE WP6 and WP1 working groups

Workshop 8 "European Platform for rare diseases registration"
Saturday 1st of June, Dubrovnik

EURORDIS
Rare Diseases Europe

EURORDIS MEMBERSHIP MEETING
DUBROVNIK 2013
30May-1June, Hotel Rixos Libertas

This presentation reports some results of the EPIRARE surveys on Rare Disease Registries

Workshop 8 "European Platform for rare diseases registration"

Saturday 1st of June, Dubrovnik

IFC CNR

Background

Two key terms and concepts

- **Rare Diseases Data Warehouse [RD-DWH]**
- **Platform for RD (registration) [PRD]**

Background

RD-DWH

is a central repository including all types of data and metadata useful for single patients, associations, health service, research scopes. *

Data can be obtained by **current statistics flows** and **ad-hoc activities** for creating reports for stakeholders and decision-makers (EU, national, regional) through the RD **web-platform**. *

Background

Usually the DWH is a web-based application, completely independent from the user's computer operating system. *

*The RD registry of Tuscany Region since 2005 has operated by a **web-platforms** for **data entry – management - output**, developed under Java and Python frameworks for the web interfaces connected with Oracle and MySQL databases.*

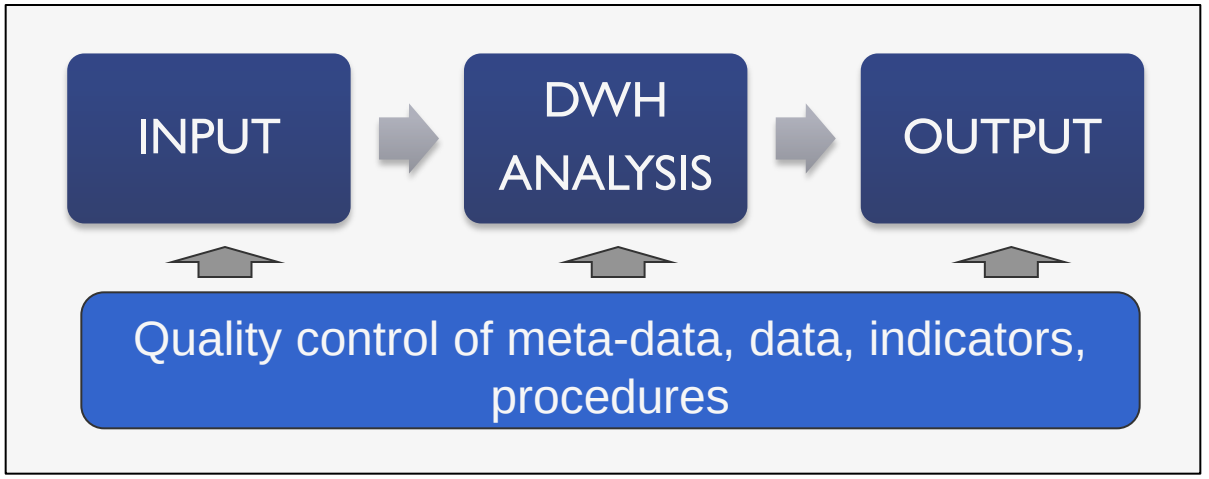
Disgression

Platform in the scientific literature, exploring pubmed

Platform → 41.566
Web platform → 1.431
Web-platform → 56
Platform AND Registry → 384
Platform AND Rare Diseases → 80
Platform AND registry AND Rare Diseases → 7
… and only few of them are relevant

Background

The data stored in DWH are uploaded from the **input module** of the **web-platform** → managed by the **analysis module** producing indicators to be transferred to → the **output module** for reporting and communication.



*

Background

INPUT

In fact, the **input module** is a set of **sub-modules** each specifically designed to "capture" the data from the different data sources and to create the **data warehouse**.

Two different procedures can be used to import data:

1. data-entry using pre-defined formats;
2. on-line, directly from the data collection sources (for example existing registries of RD, hospital discharge records, birth medical records, etc..). *

Background

INPUT

What is crucial is that:

1. the data to be imported should be specifically matched for the predefined purposes of the European RD system (EUCERD).

2. they are collected to calculate the value of the indicators, pre-defined during the **Implementation phase** that as EPIRARE we are working on.

Background

INPUT

The **import of data** is one of the crucial activities, because many of the different data set are registry/site/topic specific.

Current data (as demographics, socio-economic status, health) are calculated according to different aggregation levels, depending by the geographical aggregation availability (e.g. municipality, census tract, individual, geocoded address of residence, of hospital, of the association…).

Background

ANALYSIS

The **analysis** module, the engine of the platform, consists of a set of masks and filters designed to query different data sets for the purpose of carrying out pre-definite univariate, bivariate and multivariate analyses.

Those analyses calculate the indicators defined during the **Implementation Phase.**

Background

OUTPUT

The **output module** allows the users to produce tailored recommendations, to inform decision-making in the field of health planning and management (treatment, prevention, social aids, services planning, research), according to the EC goals. *

The EPIRARE – EC DG-SANCO funded project is working to enhance knowledge on the above summarised activities.

Two EPIRARE surveys, on 220 RDR and 180 out of 219, were carried out in order to investigate the existing state-of-art of RD registration in European countries.

Introduction

- The challenge will be to identify a common data elements, defined consistently with aim of Registry, that has to be able to standardize the data collection of the rare diseases.

- Analyses were performed to understand the needs and the informative abilities of currently existing Registries, in order to provide a common and shareable informative platform.

Materials

Data from the Survey EPIRARE were analyzed, regarding:

- ✓Aims
- ✓Population Target
- ✓Number of diseases
- ✓Data providers
- ✓Type of data collected
- ✓Disease Coding System
- ✓Data sharing

Methods

- Univariate analysis of response mode
- Potential associations among variables were investigated by multivariate analysis using Logistic Regression models
- Factor analysis was performed and was mainly oriented to find out the structure of latent relationships among variables, using the Multiple Correspondence Analysis

Distribution by countries of the 220 RDR responding at the first EPIRARE survey's questionnaire

Country	Number	%
France	27	12,3
Germany	20	9,1
Italy	29	13,2
Spain	46	20,9
UK	16	7,3
Other European Countries	69	31,4
Not European Countries	13	5,9

Aims

Aim	N	%
Epidemiological research	155	70,8
Clinical research	134	61,2
Natural history of disease	133	60,7
Disease surveillance	122	55,7
Genotype-phenotype correlation	117	53,4
Mutation database	94	42,9
Treatment evaluation (efficacy)	94	42,9
Healthcare service planning	74	33,8
Treatment monitoring (safety)	73	33,3
Social planning	42	19,2
Other	18	8,2

Target Population

Target	N	%
Population-based	124	57,1
Hospital-based	52	24,0
Case-based*	41	18,9

** Case series + Family case series + Cohort of cases*

Number of rare diseases included

N. Rare disease	N	%
Single RD	75	34,3
A group of related RDs	102	46,6
Several RDs (not related)	26	11,9
All rare disease	16	7,3

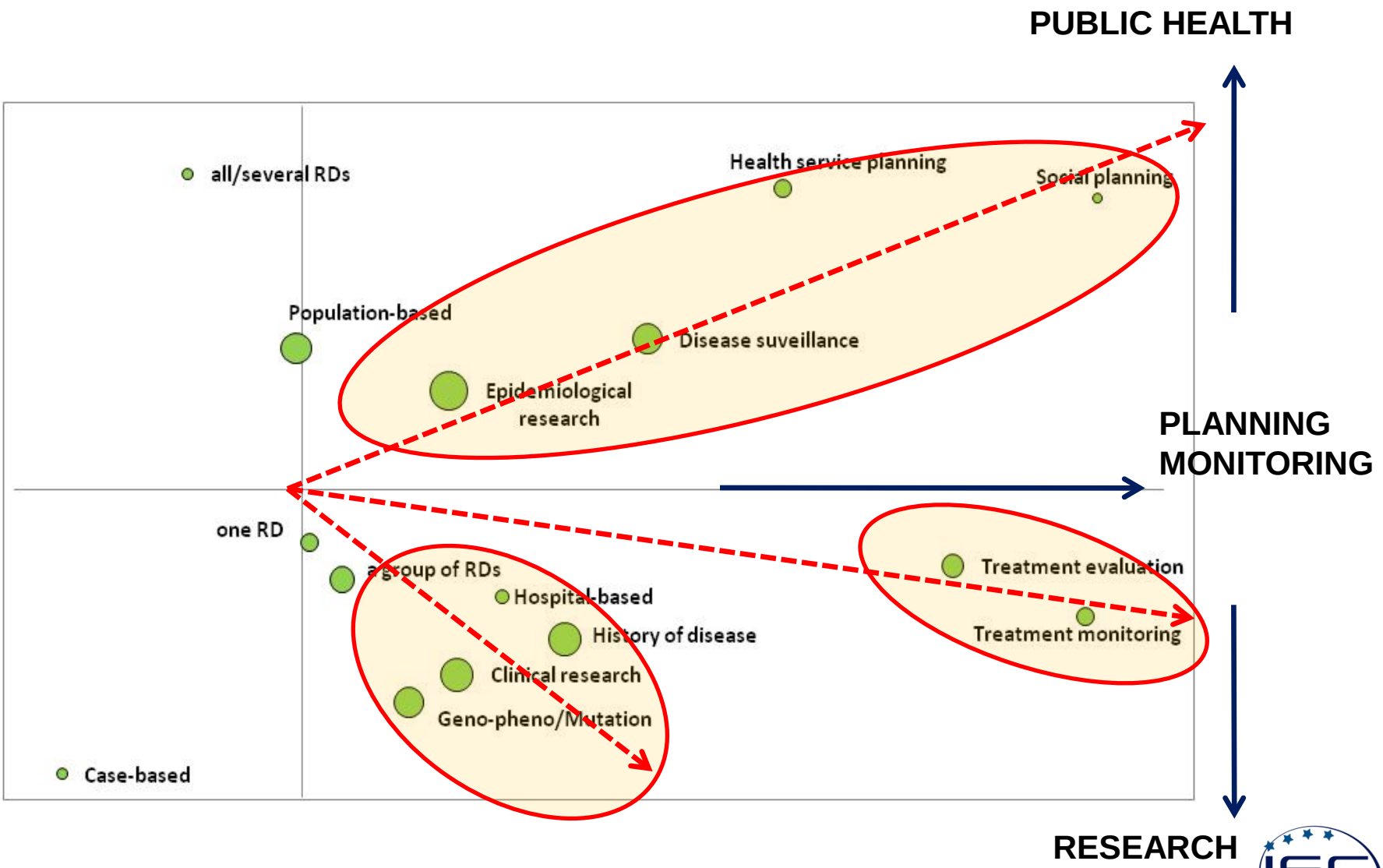
Exploratory Statistical Analysis on 1st Survey data

two types of statistical analysis were done:

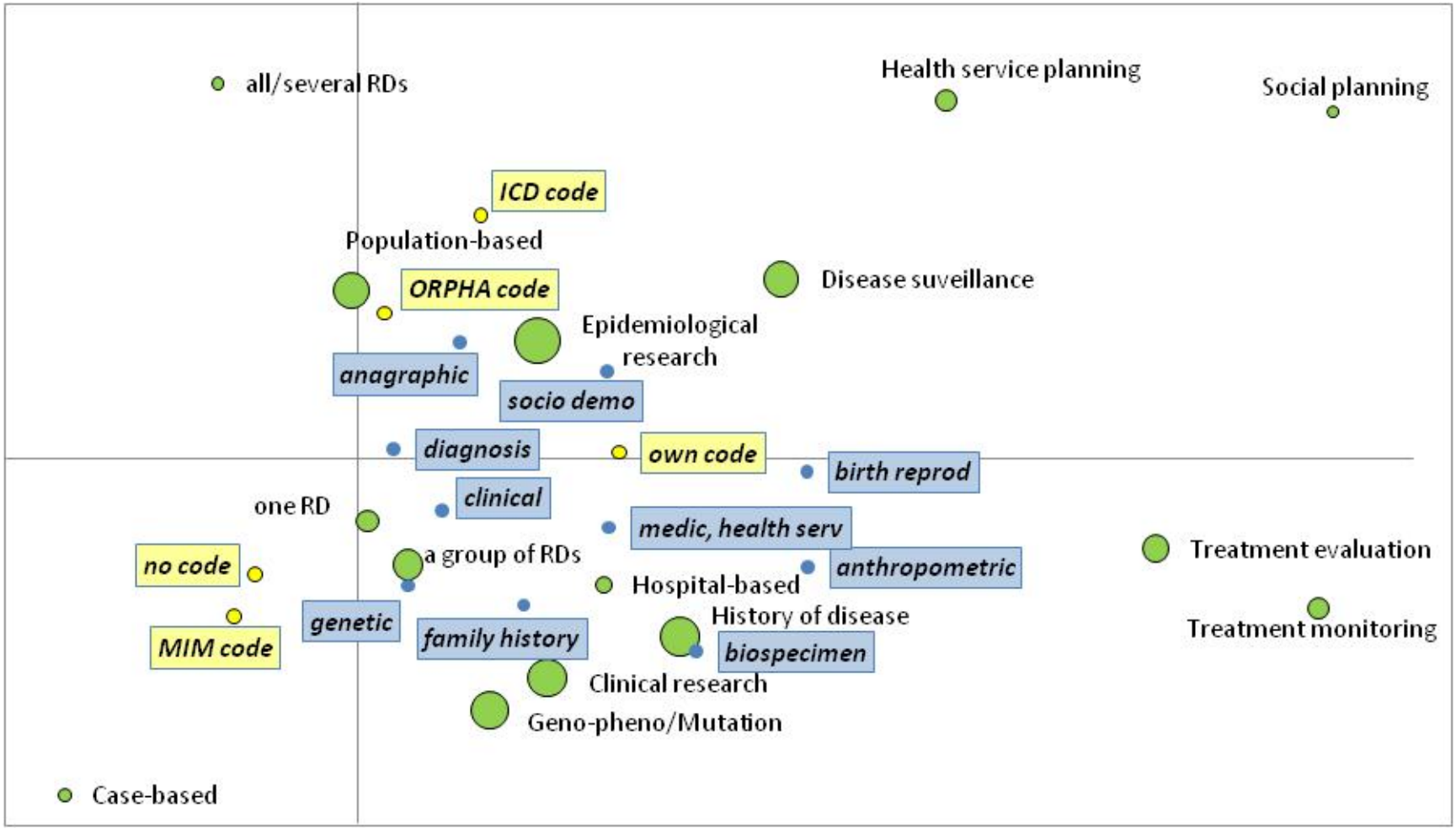
Correspondence Analysis → *Variables*

Cluster Analysis → *Observations (Registries)*

MCA: Factorial Plan by the active variables

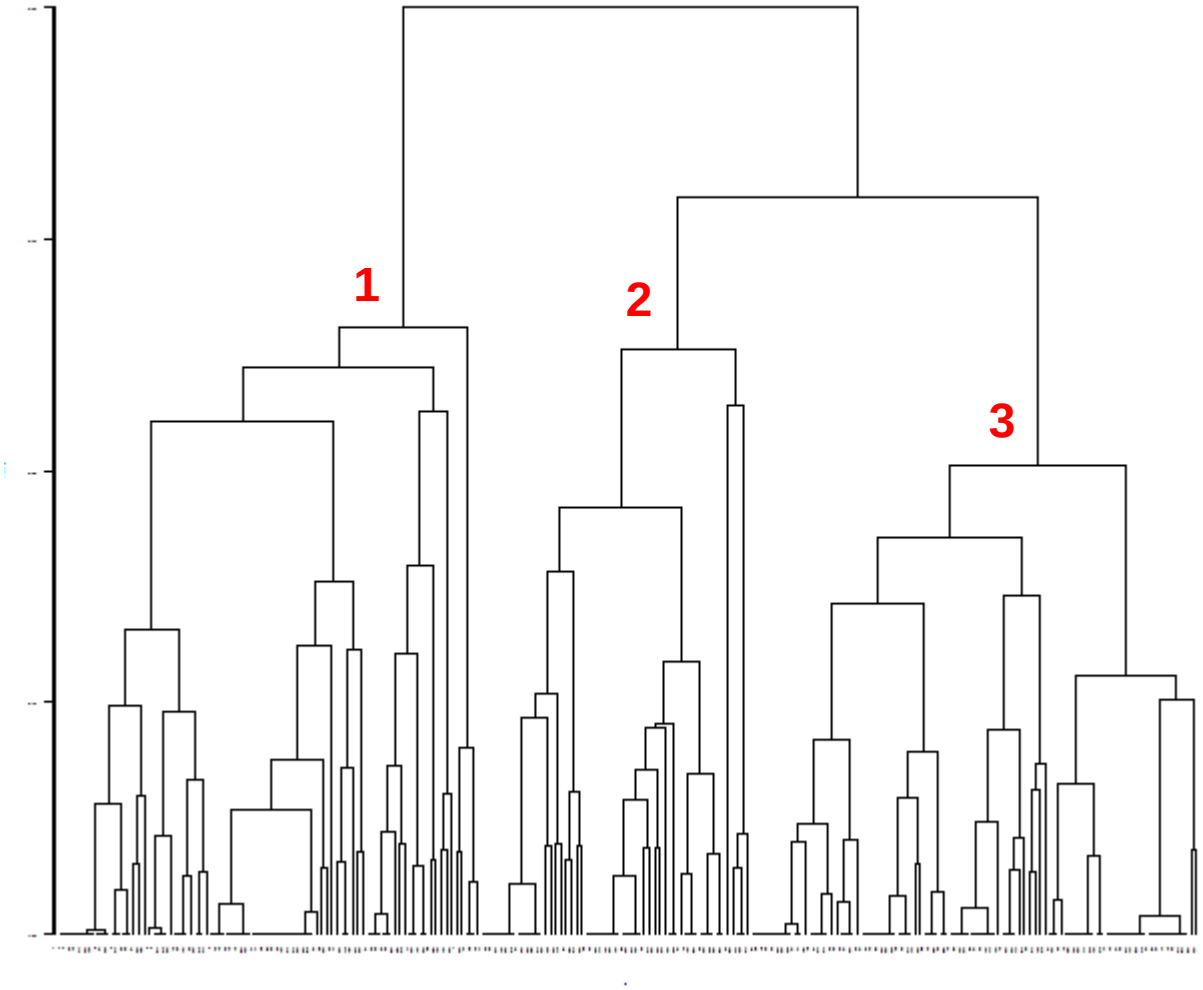


MCA: Projection of "Data collected" and "Disease Coding System"



Cluster Analysis

Tree Diagram



Distribution of the Aims in the Clusters

Aim	CL 1	CL 2	CL 3
	%	%	%
Epidemiological Research	90.4	55.8	74.1
Disease surveillance	75.0	25.6	75.3
Healthcare service planning	63.5	7.0	43.2
Social planning	32.7	2.3	28.4
Clinical Research	21.2	69.8	77.8
Natural history of disease	11.5	66.3	86.4
Mutation database / Genotype-phenotype correlation	3.8	80.2	69.1
Treatment evaluation	17.3	5.8	98.8
Treatment monitoring	5.8	4.7	81.5

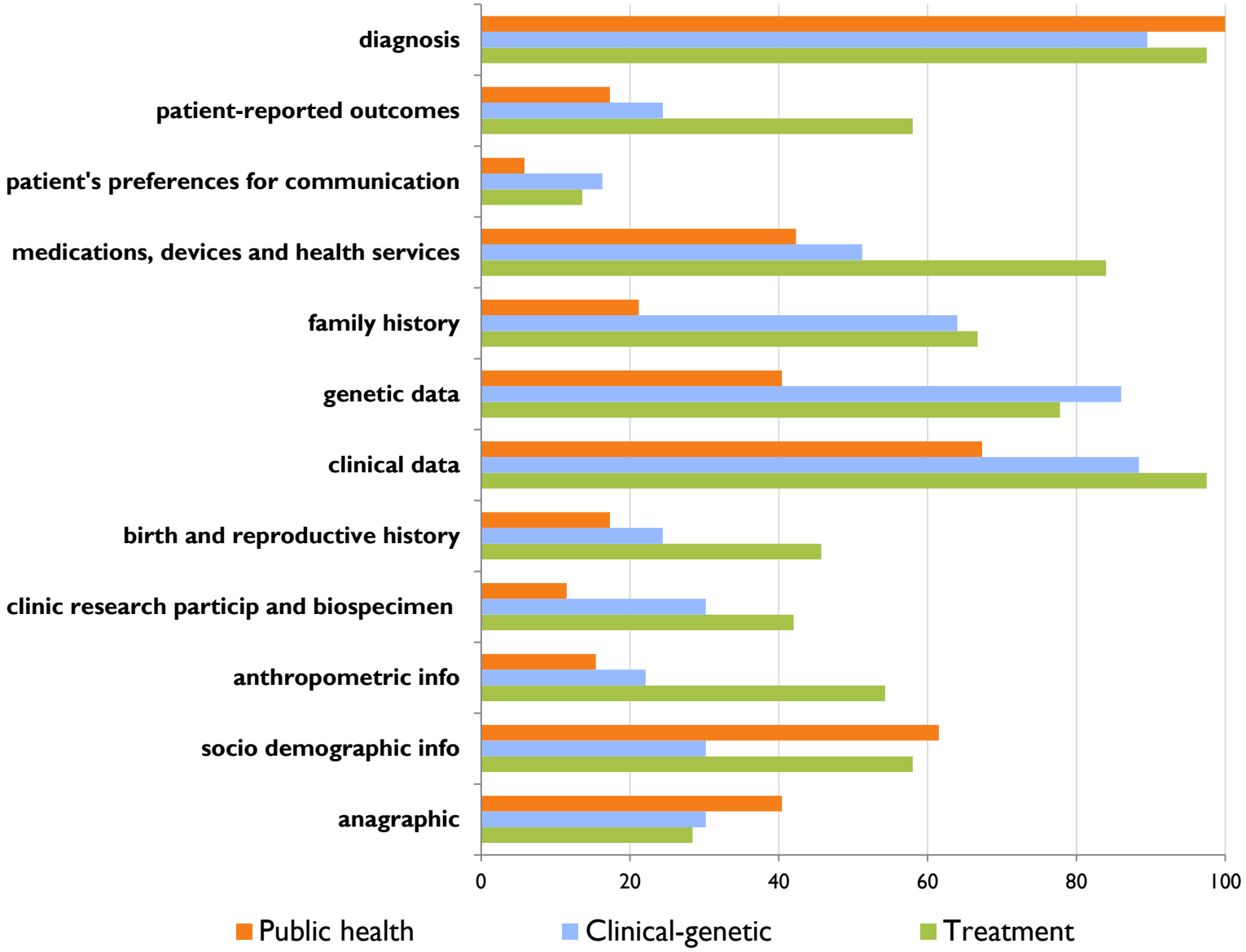
Interpretation and definition of the Clusters

Cluster 1: Public Health

Cluster 2: Clinical and Genetic Research

Cluster 3: Treatment

Data collected



SURVEY EPIRARE DATA ELEMENTS – Analysis of WP6

Objective of Analysis:

- **feasibility of collection specific Data Elements**

- **different capability of the types of Registries to collect specific Data Elements**

SURVEY EPIRARE DATA ELEMENTS – Analysis of WP6

Classification of respondent Registries based on Aims, number of diseases, network

Identification of 3 main types of Registries:

1.Public Health

2.Clinical Research without activity on Treatment

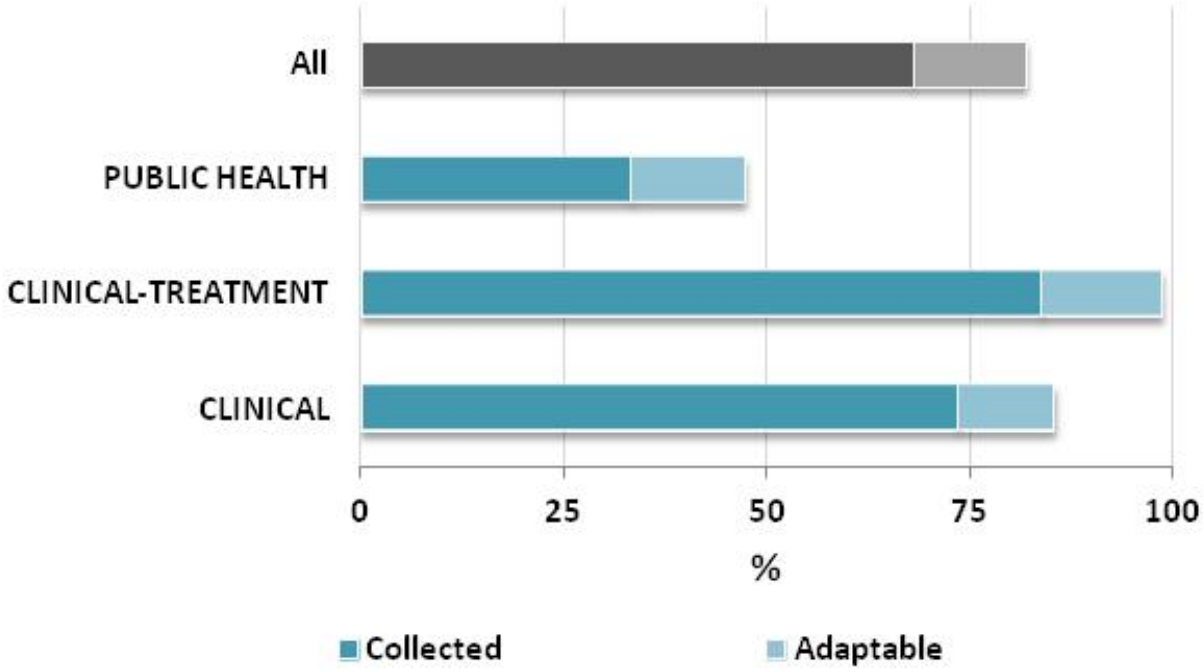
3.Clinical Research with activities on Treatment

Classification of the Registries

Type of Registry	N	%
PUBLIC HEALTH	36	25.7
CLINICAL	36	25.7
CLINICAL-TREATMENT	68	48.6

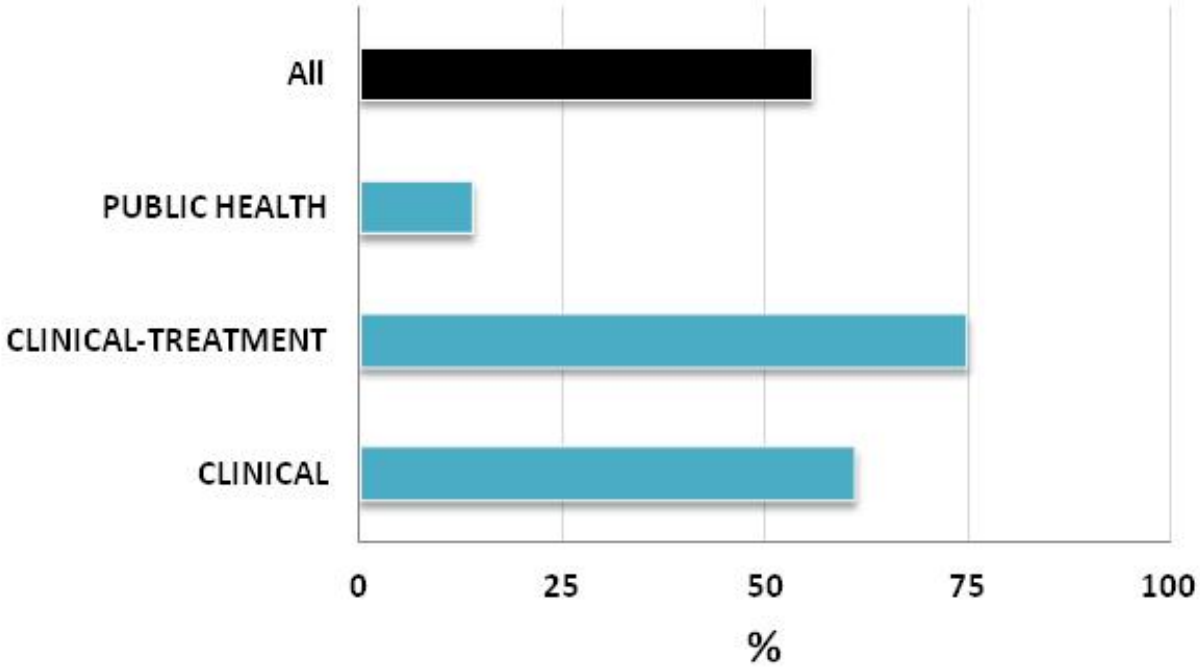
Goodness of classification of the Registries

Could your Registry collect the Clinical data of disease?

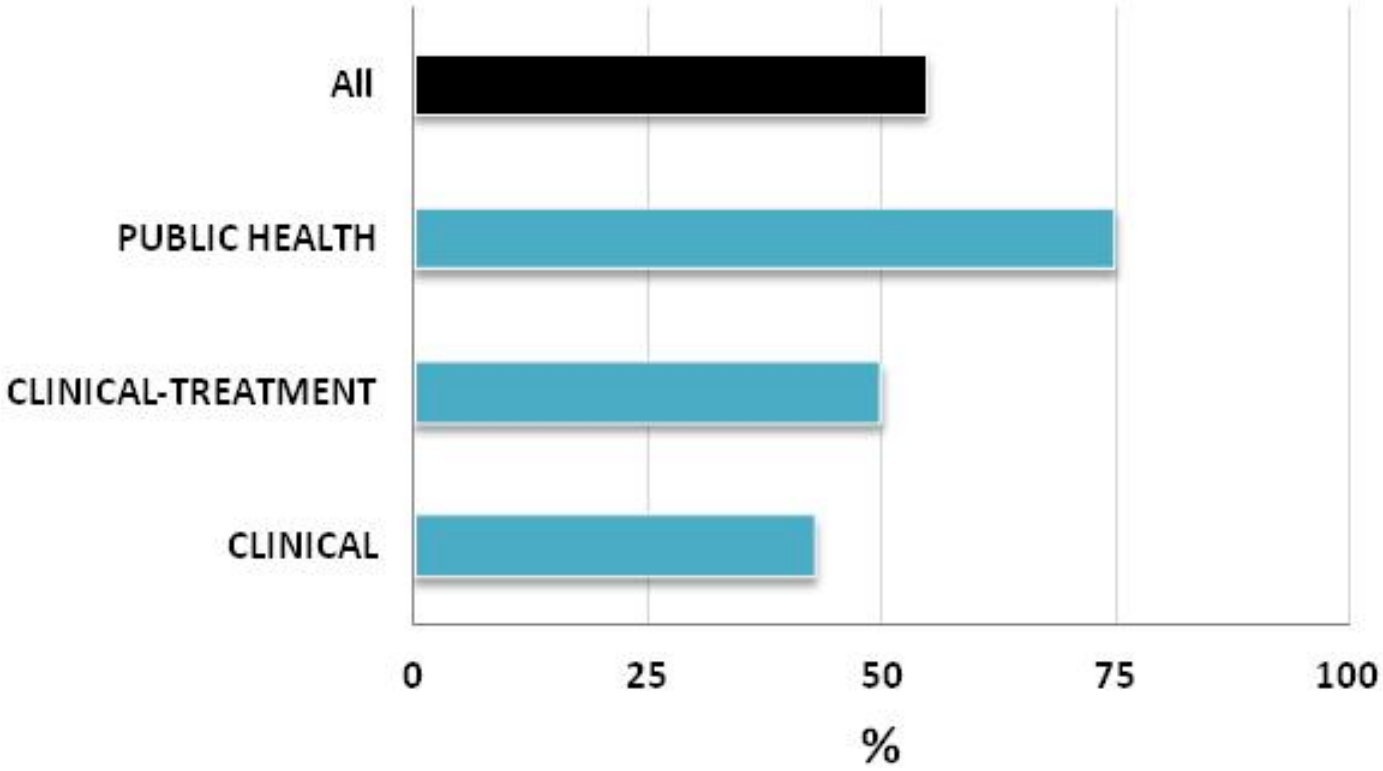


Goodness of classification of the Registries

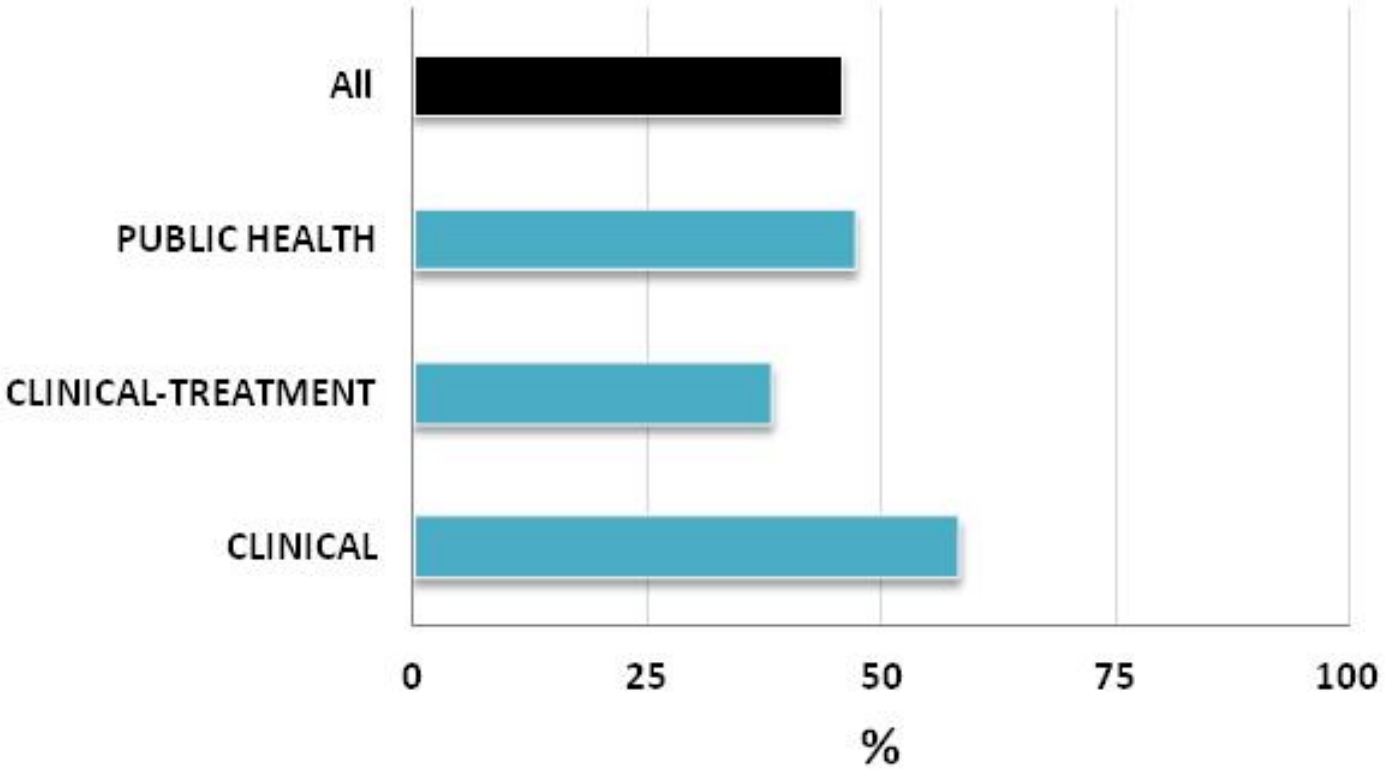
Registries declaring as Aim: <u>Genotype-phenotype correlation/Mutation database</u>



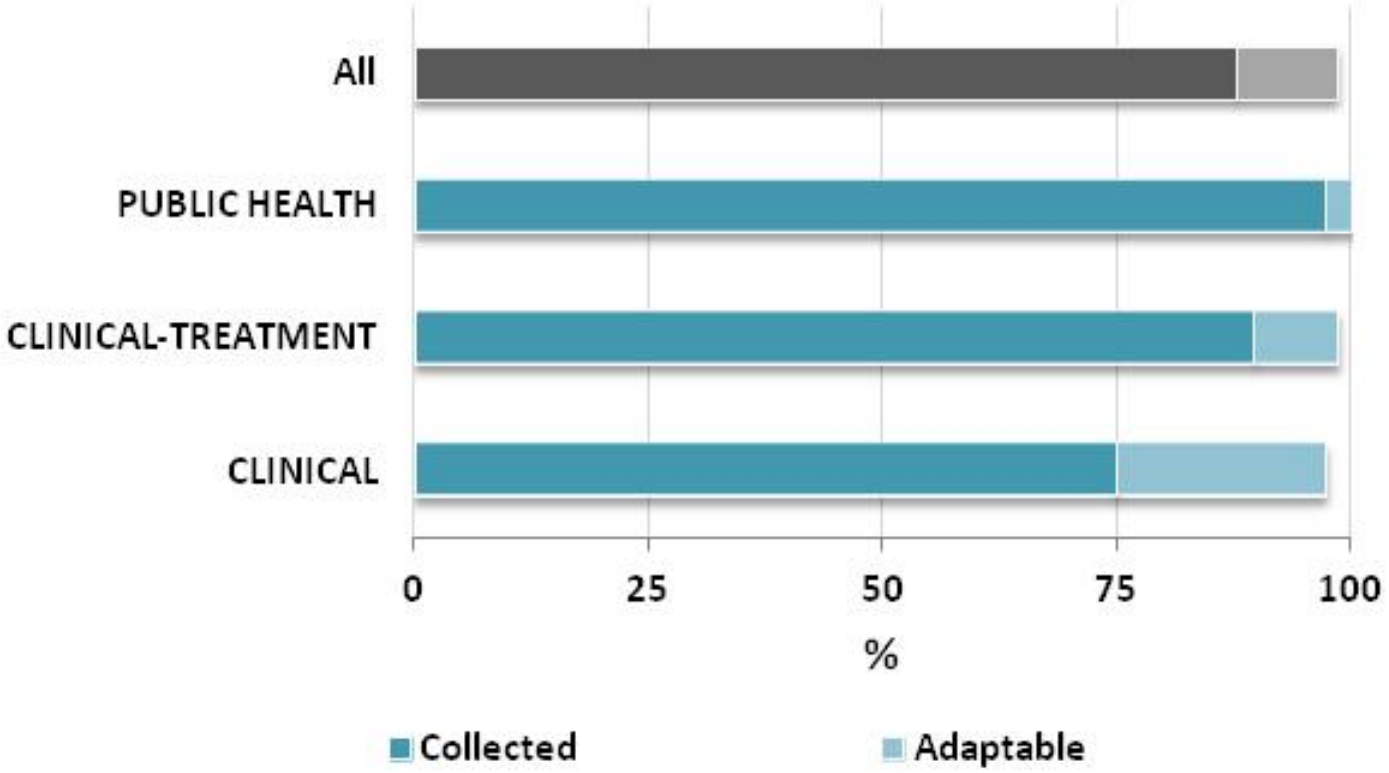
Registry collects data on Patients with confirmed diagnosis only ?



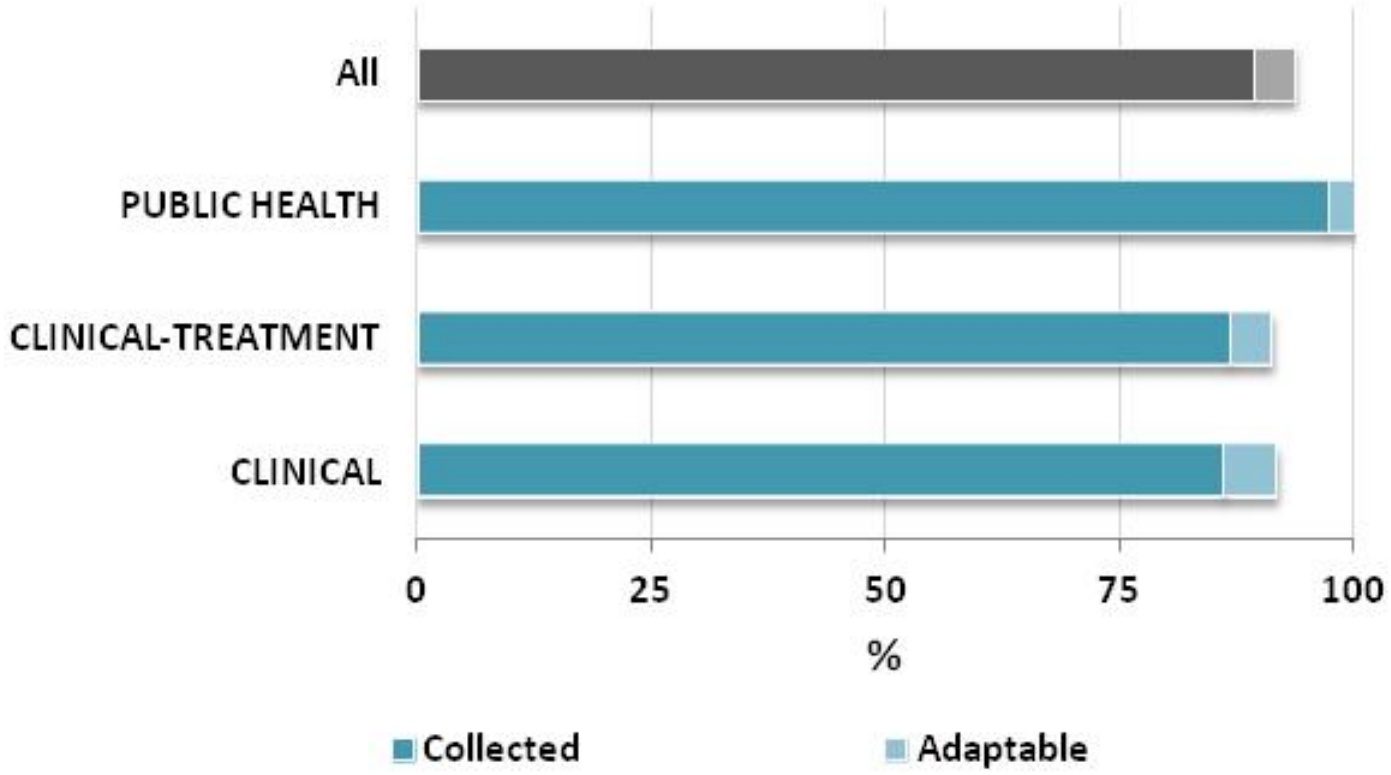
***Key connection**: First Name + Last name + Sex + Date of birth + City of birth + Country of birth*



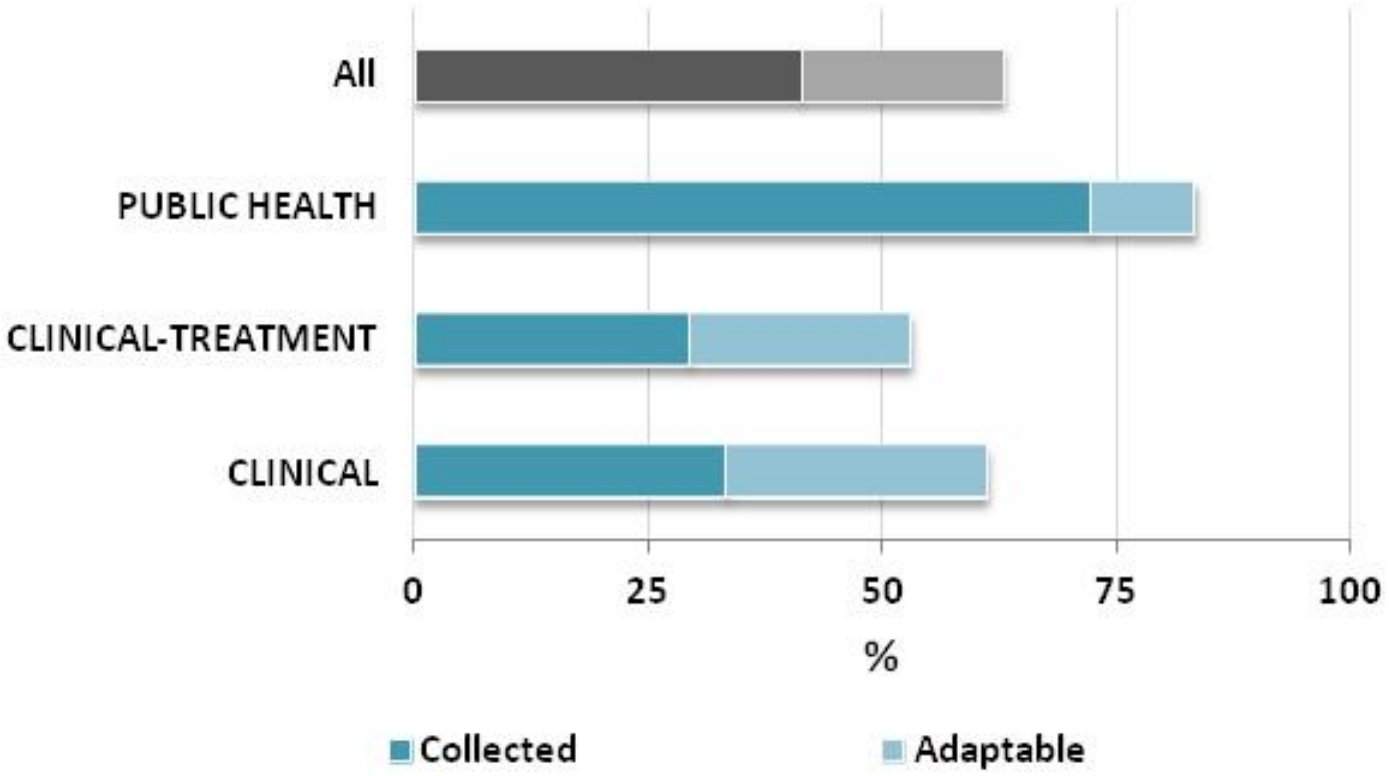
Could your Registry collect the Gender of patient?



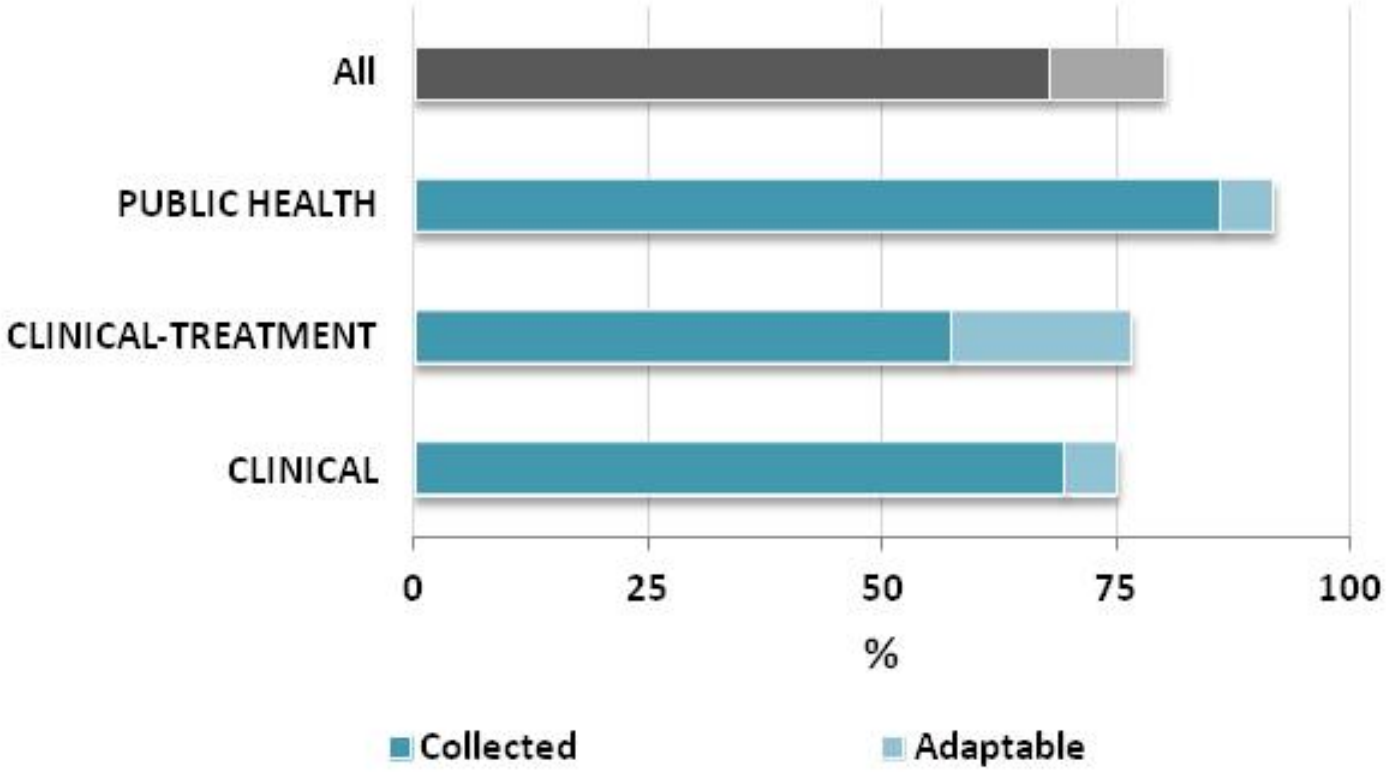
Could your Registry collect the Date of birth of patient?



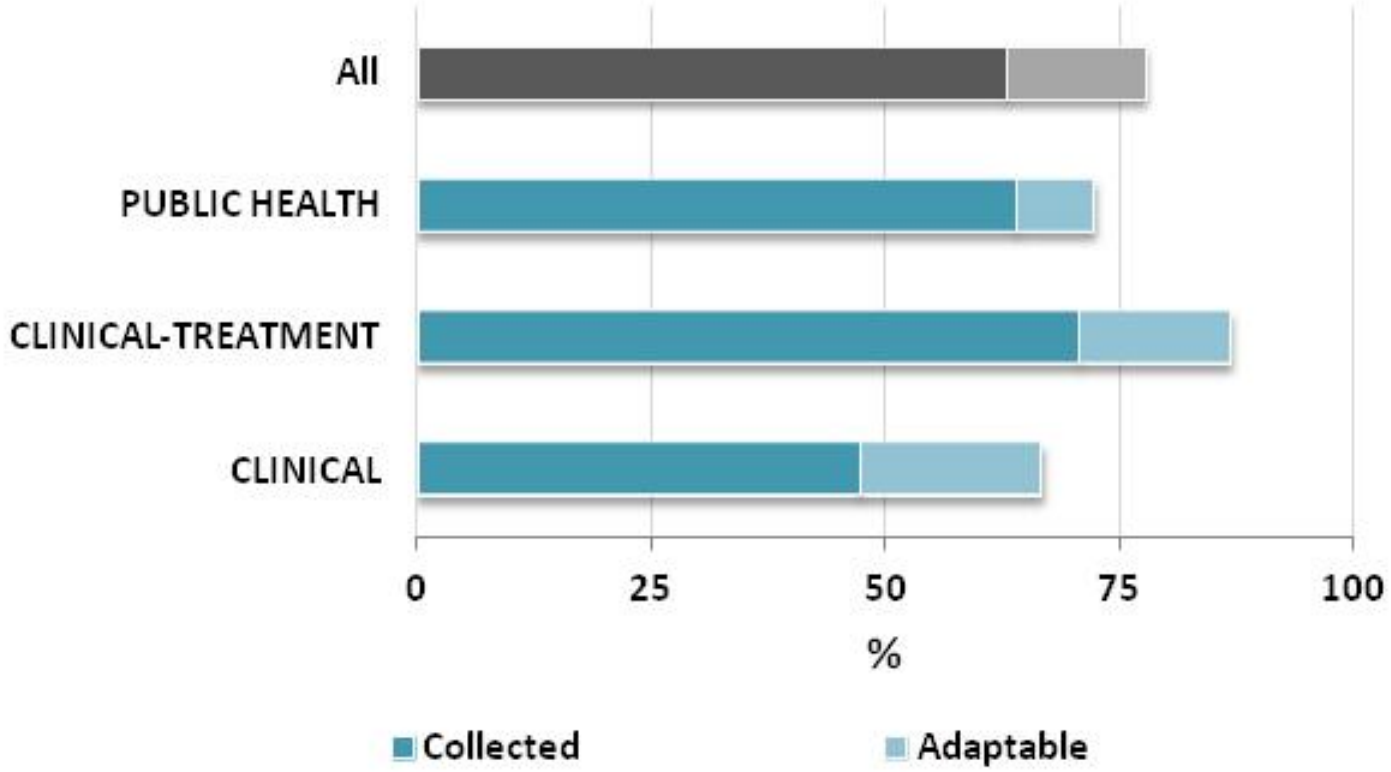
Could your Registry collect the Patient's city of birth?



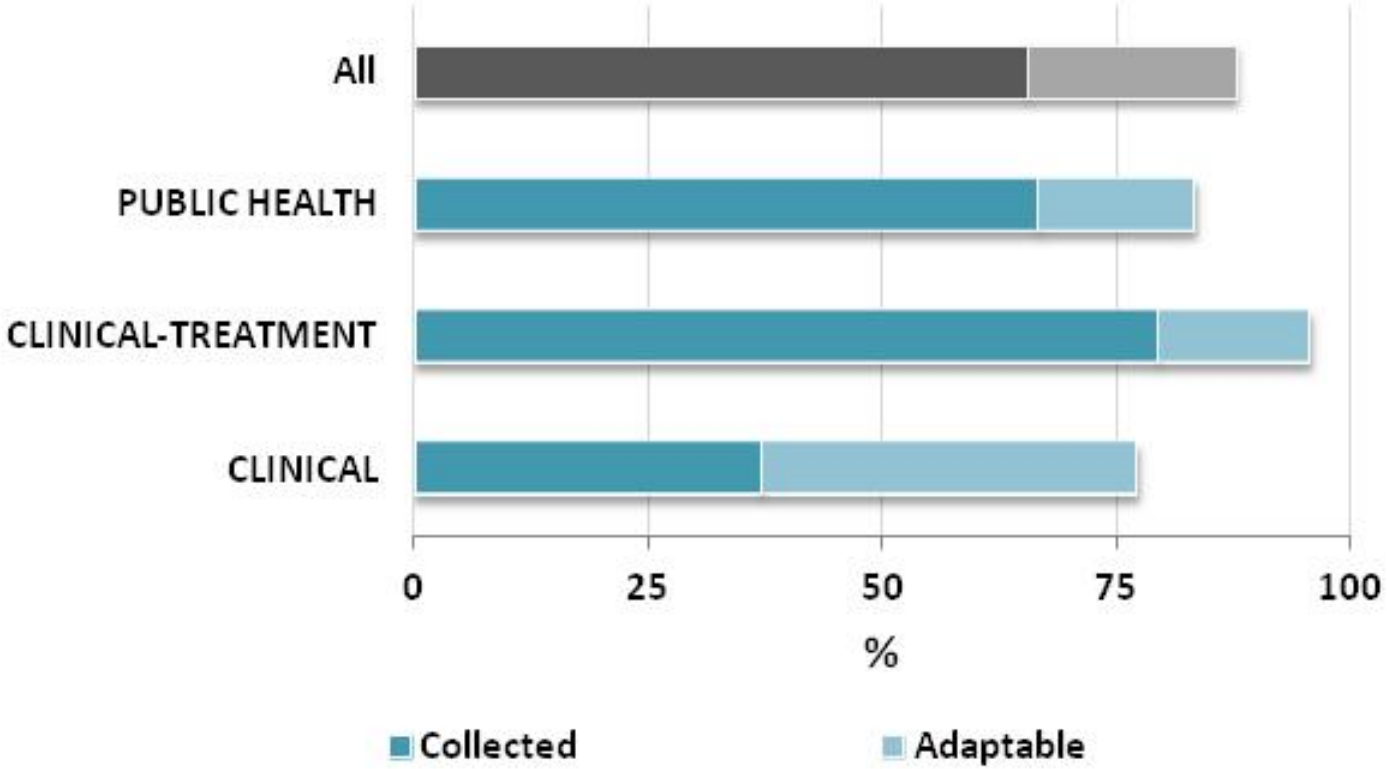
Could your Registry collect the Patient's city of residence?



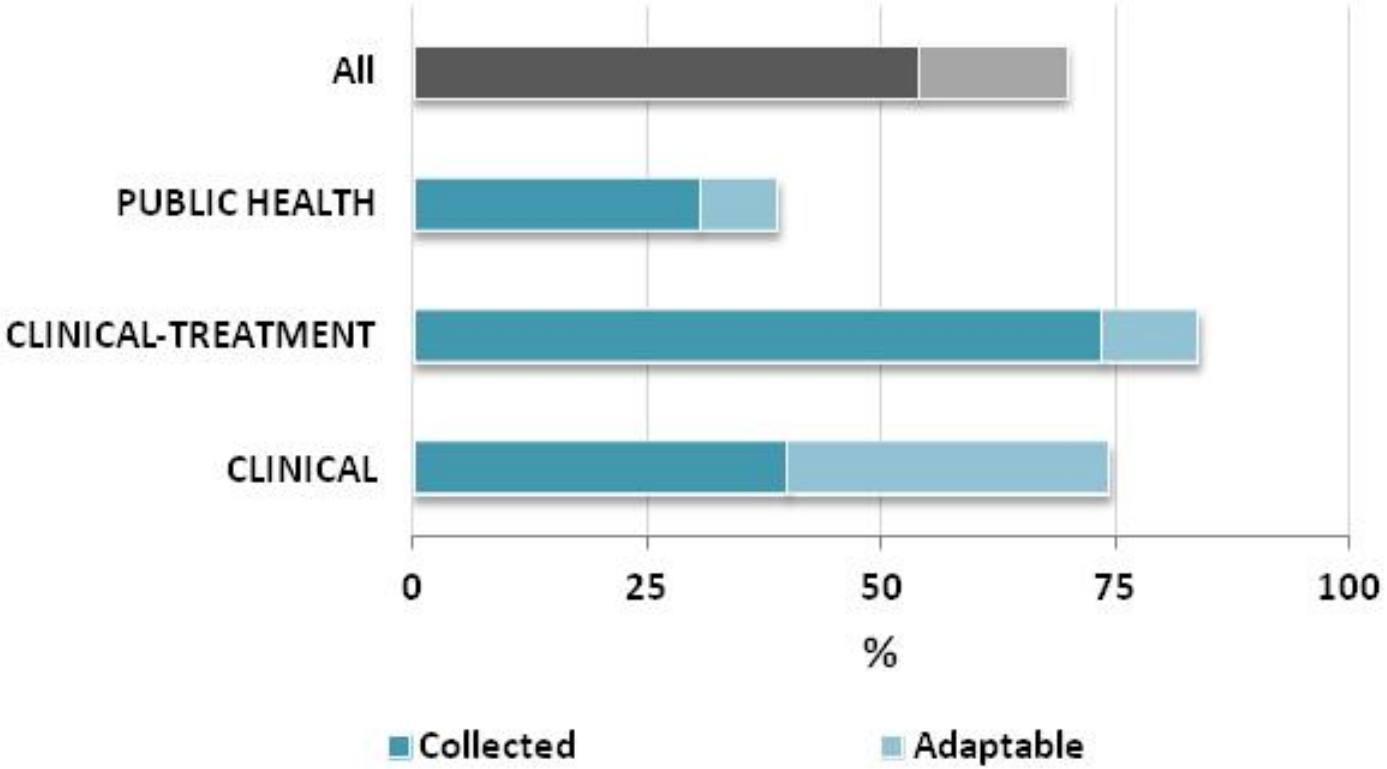
Could your Registry collect the Patient's date of death?



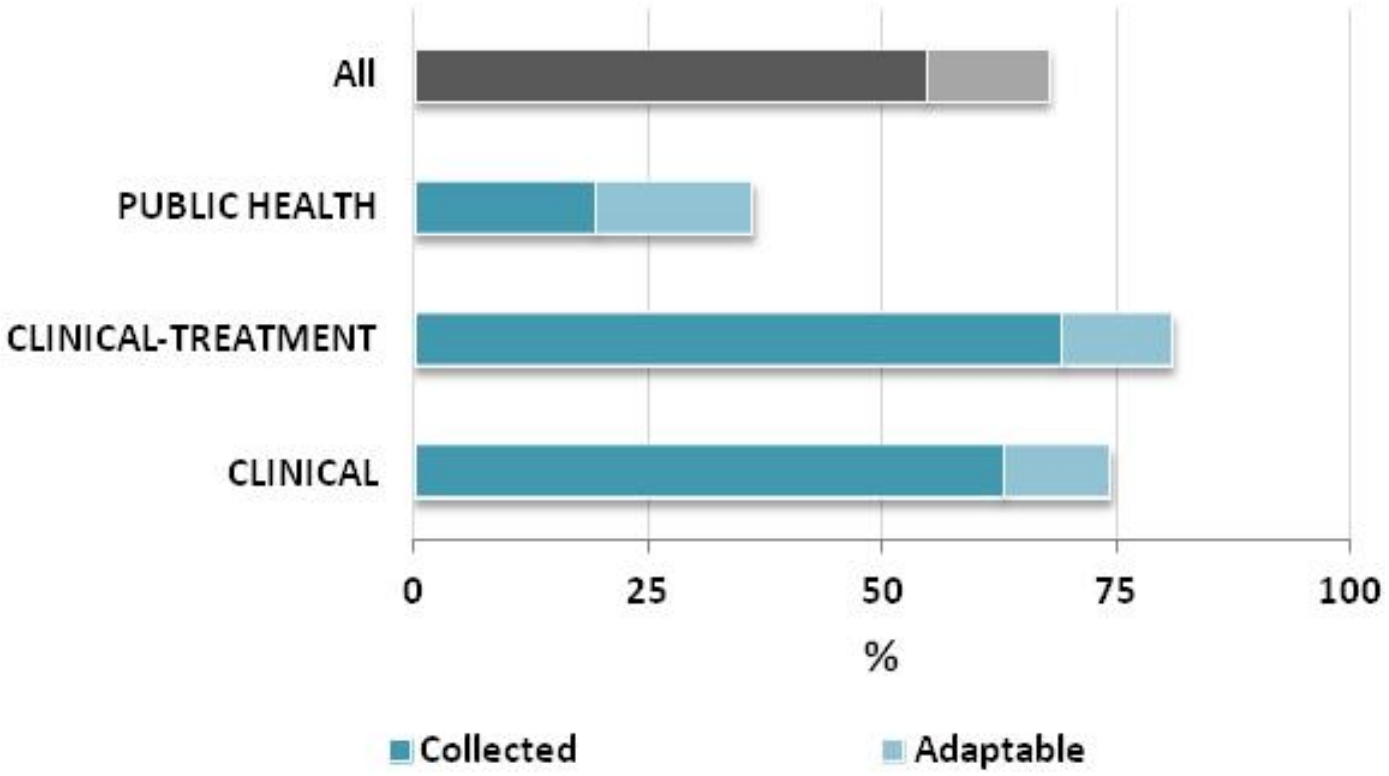
Could your Registry collect the Date of diagnosis?



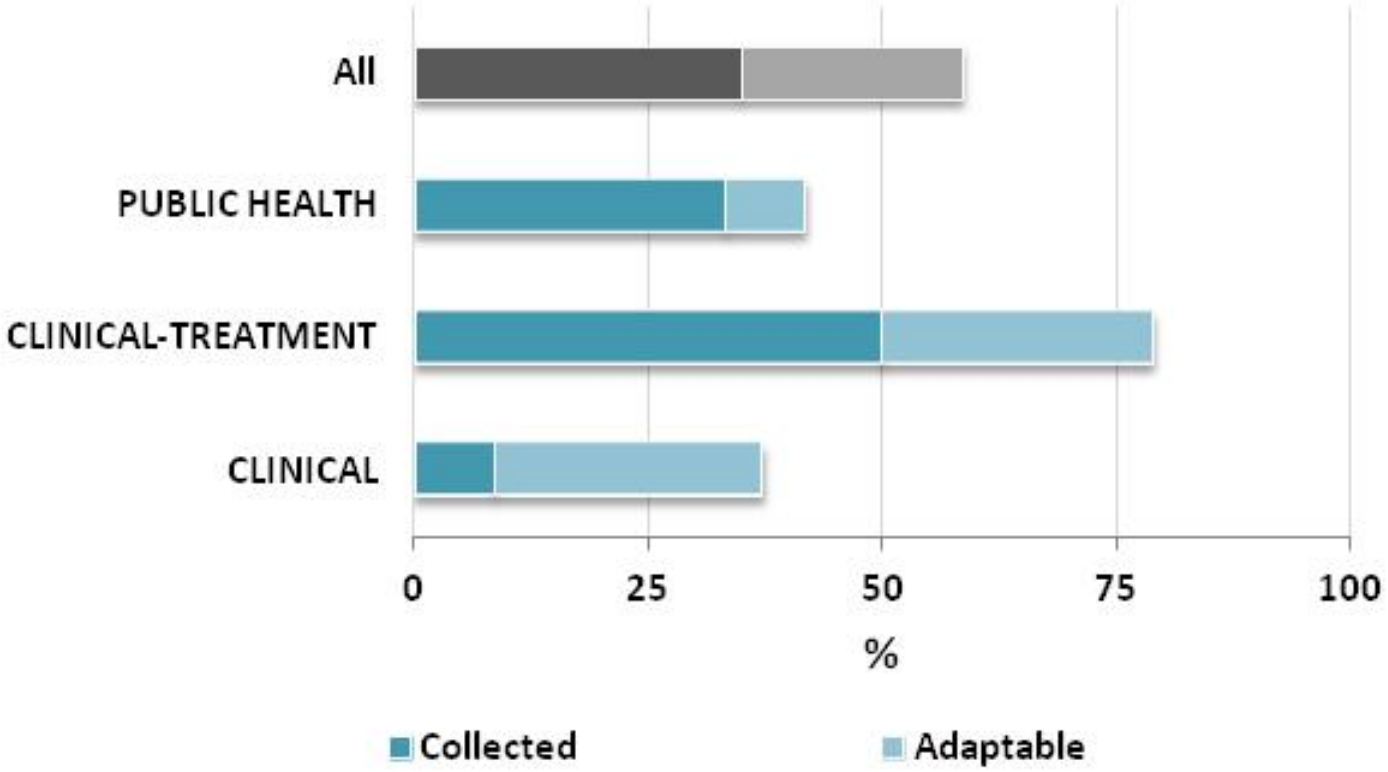
Could your Registry collect the Patient's date of disease onset?



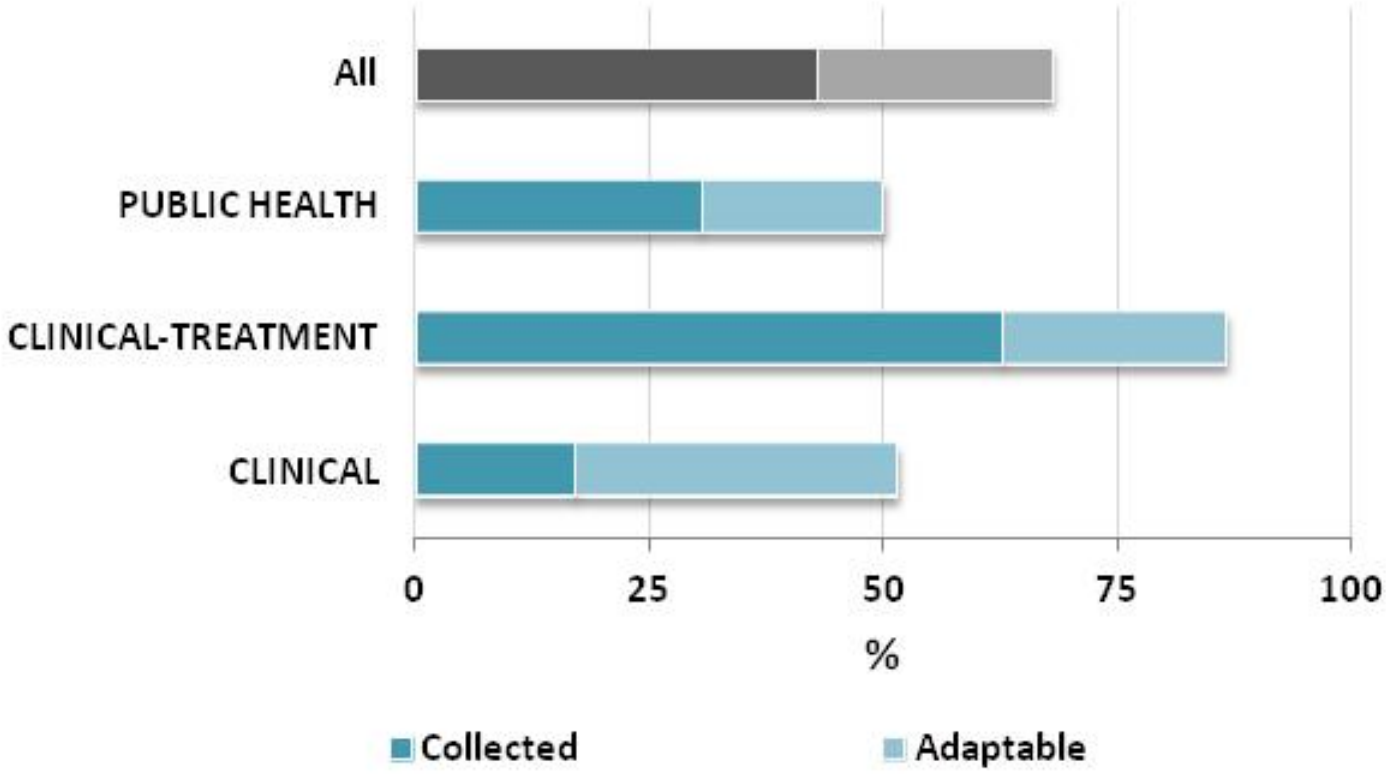
Could your Registry collect the Co-morbidity?



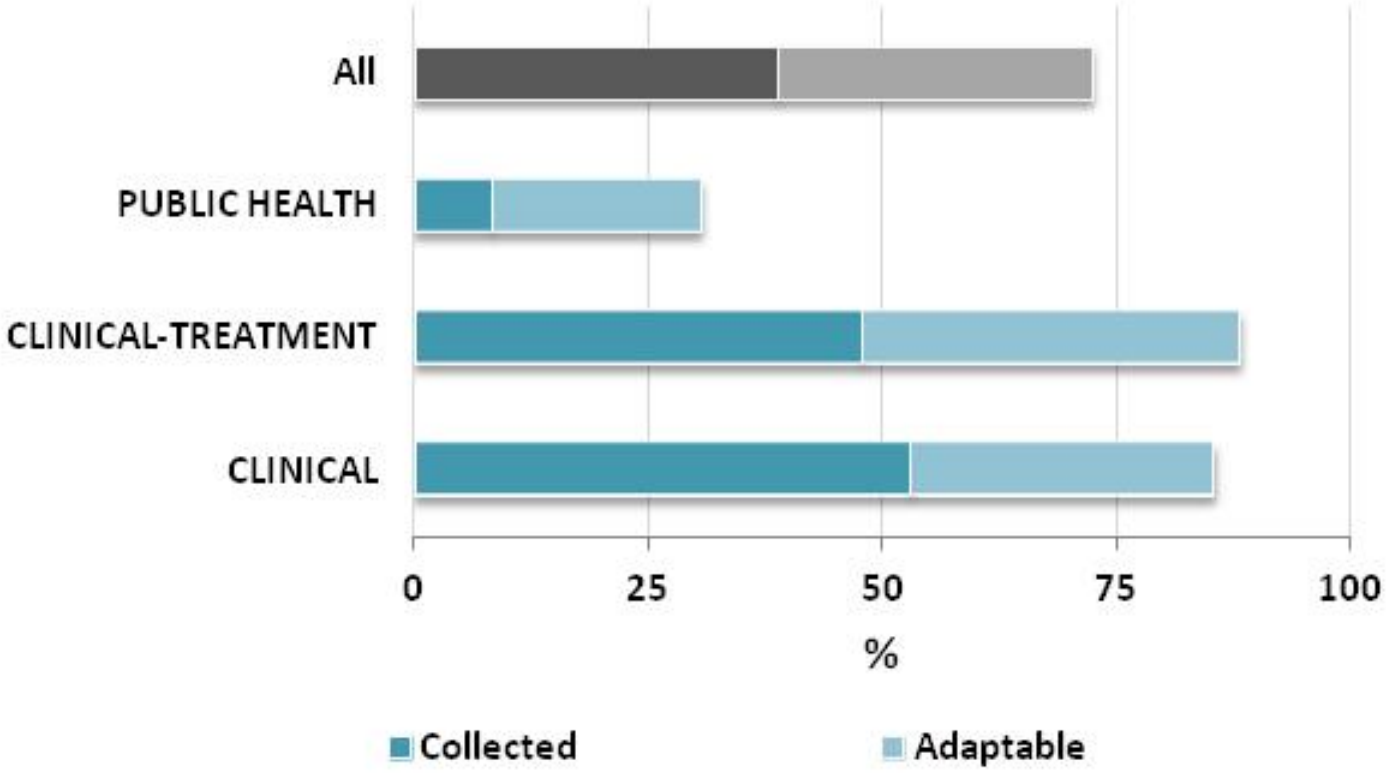
Could your Registry collect data on Current orphan drug treatment?



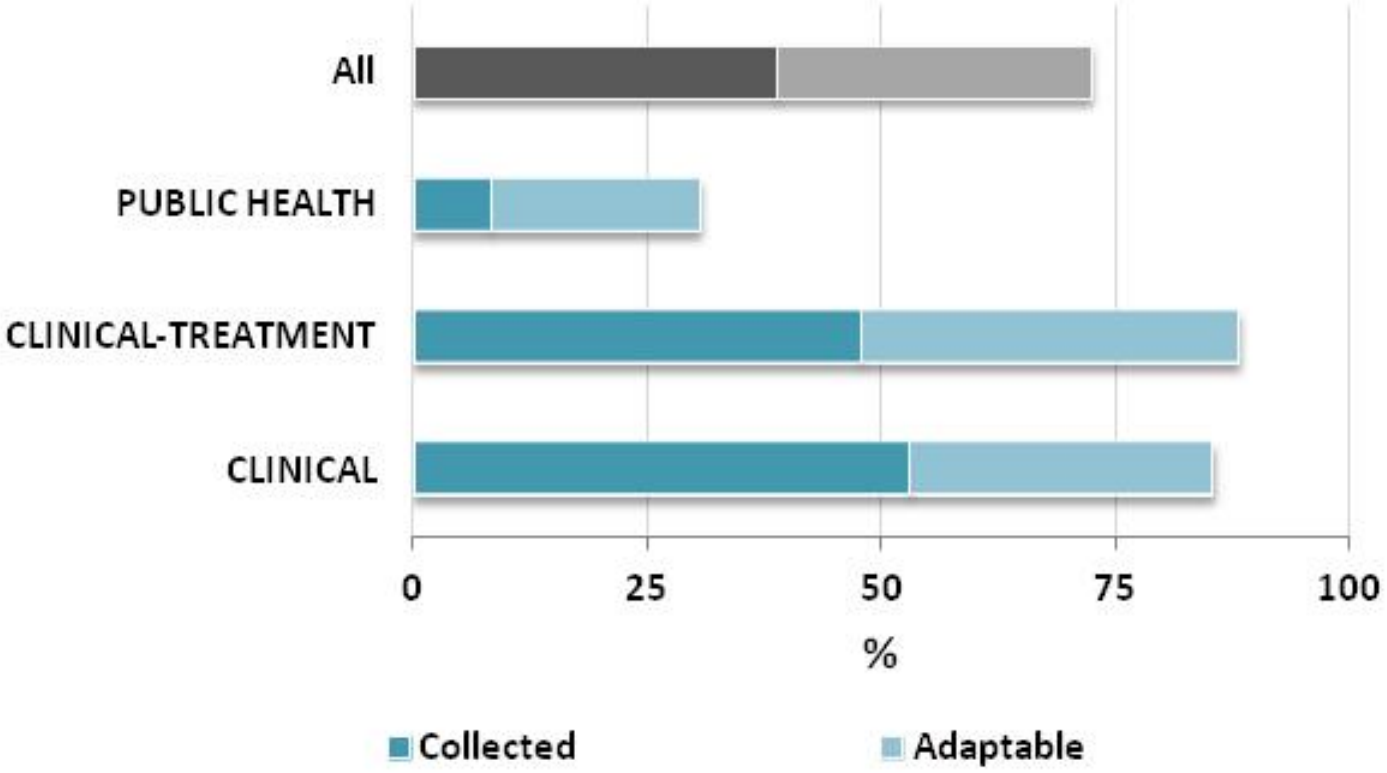
Could your Registry collect data on Current treatment?



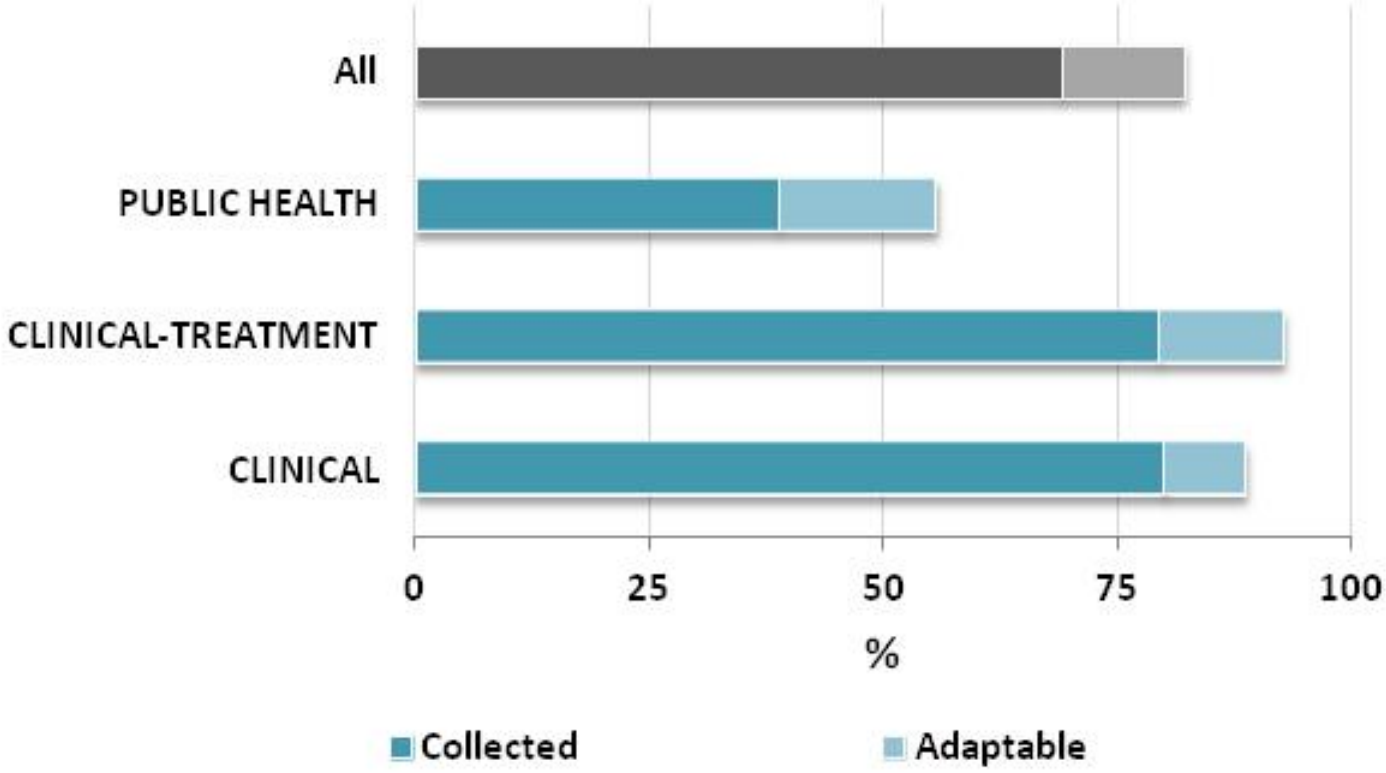
Patient's willingness to be contacted to participate in a future trial



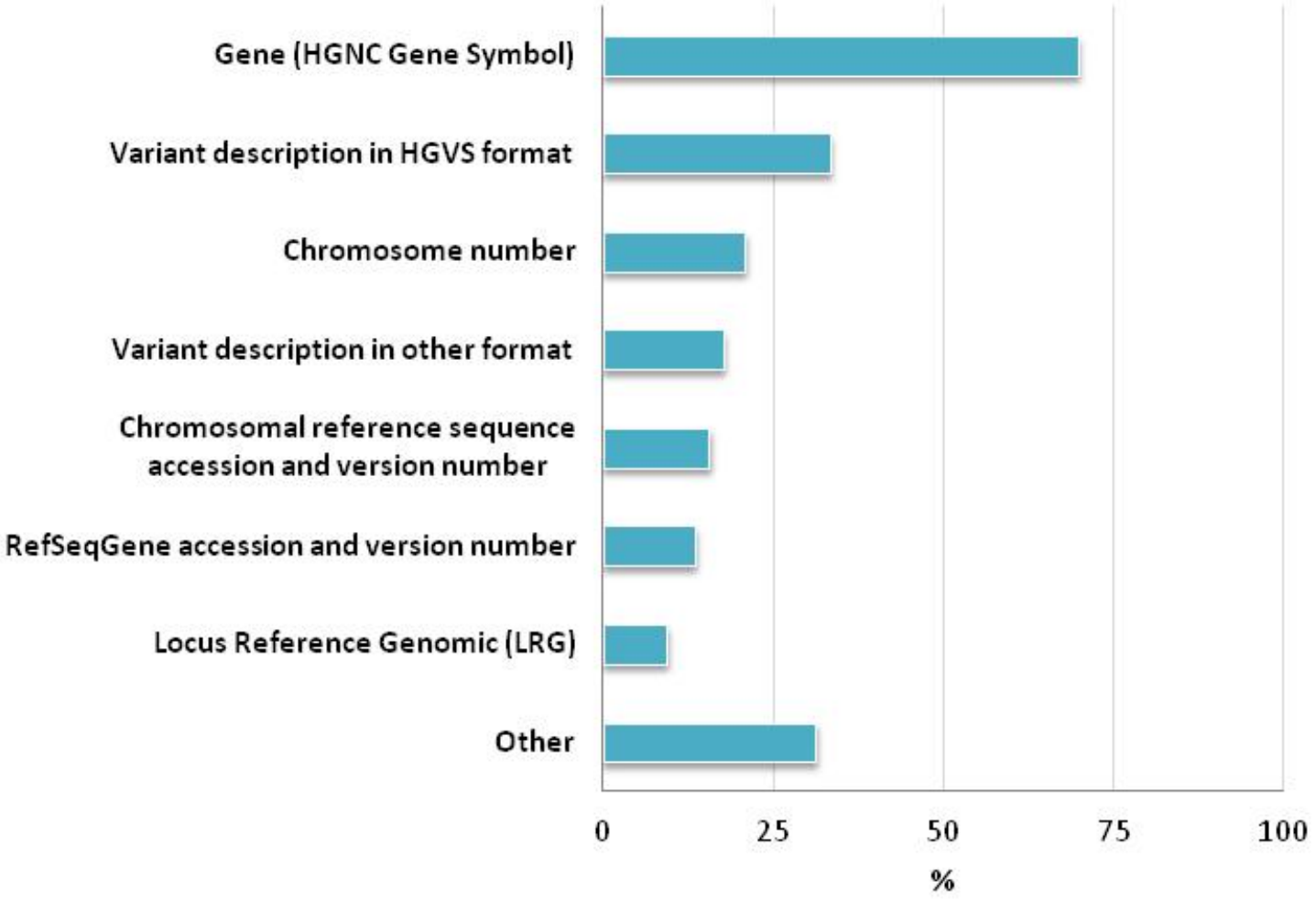
Patient's willingness to be contacted about donating biological samples for research



Is your Registry able to collect data on Genetic features of the rare disease?



If collected data on Genetic features, what genetic data?



Conclusions

The analysis identifies registries

- Population oriented

- Disease oriented

AND

- Public health

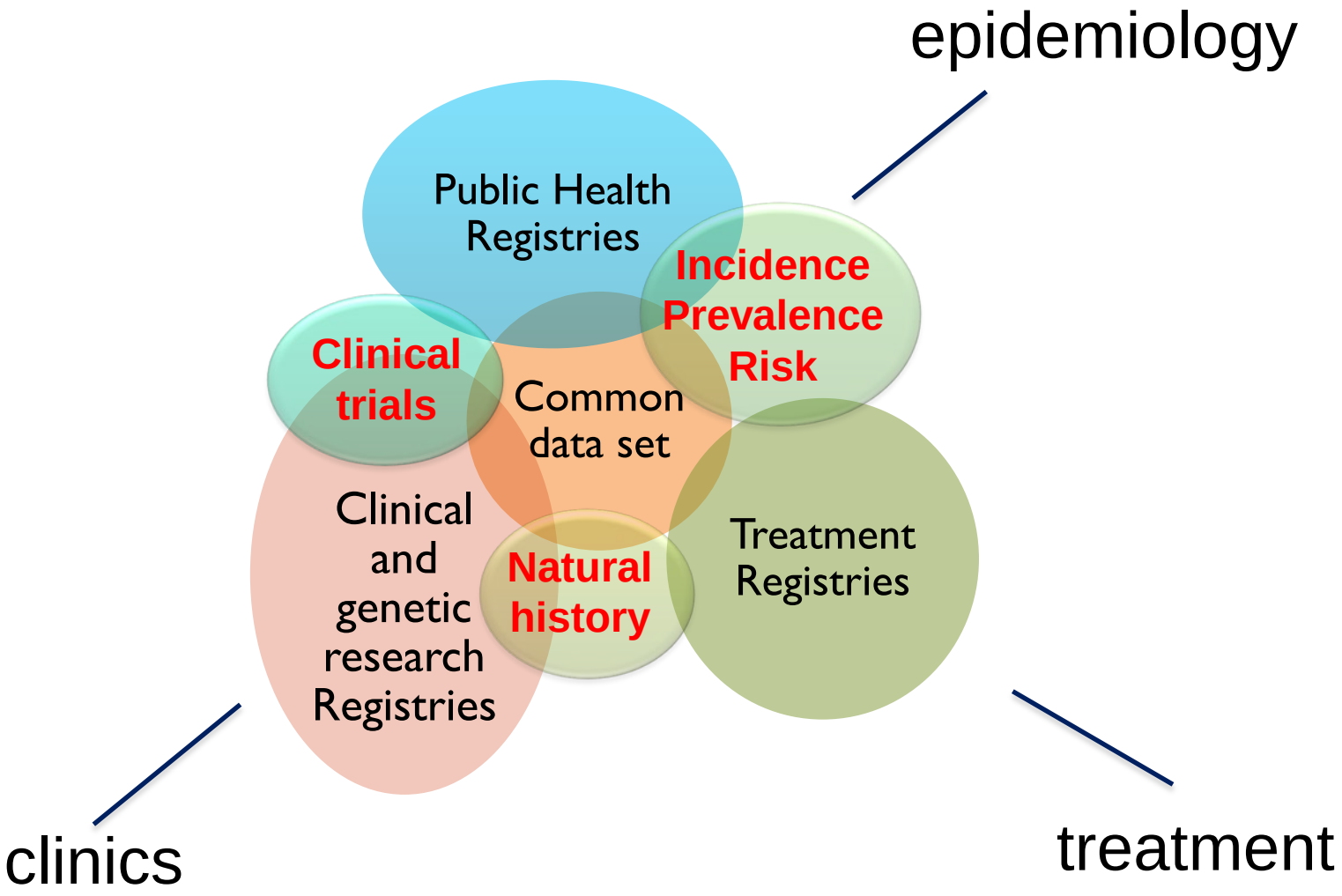
- Clinical with or without treatment

having different data but also a common body of information

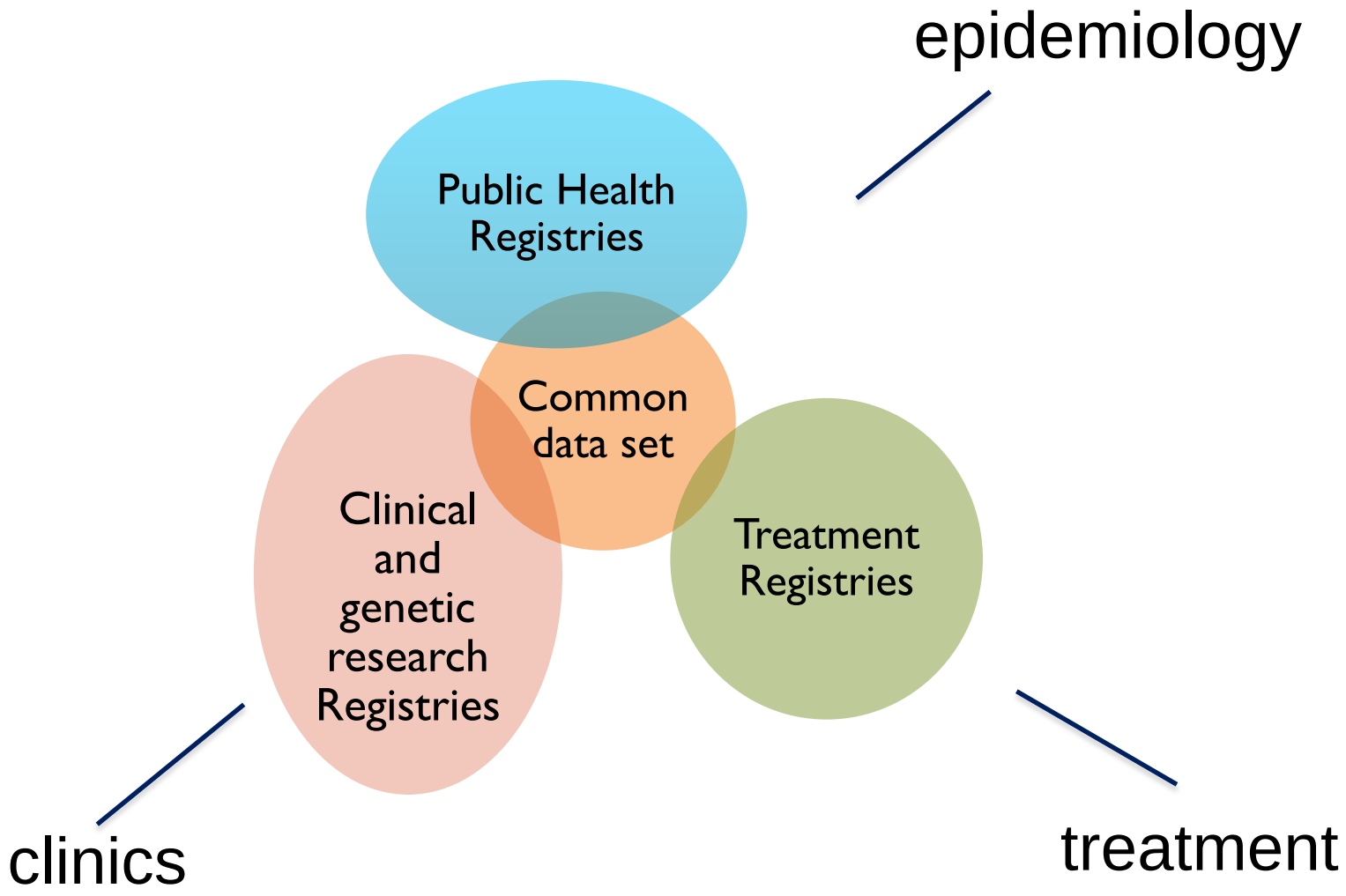
Differences and common traits are under study to:

- define a common data set for all RDs registries
- Identify a prototype of RDs platform, i.e. a flexible object with a common body and many satellites, able to include different registries for orienting Research and Public health activities. *

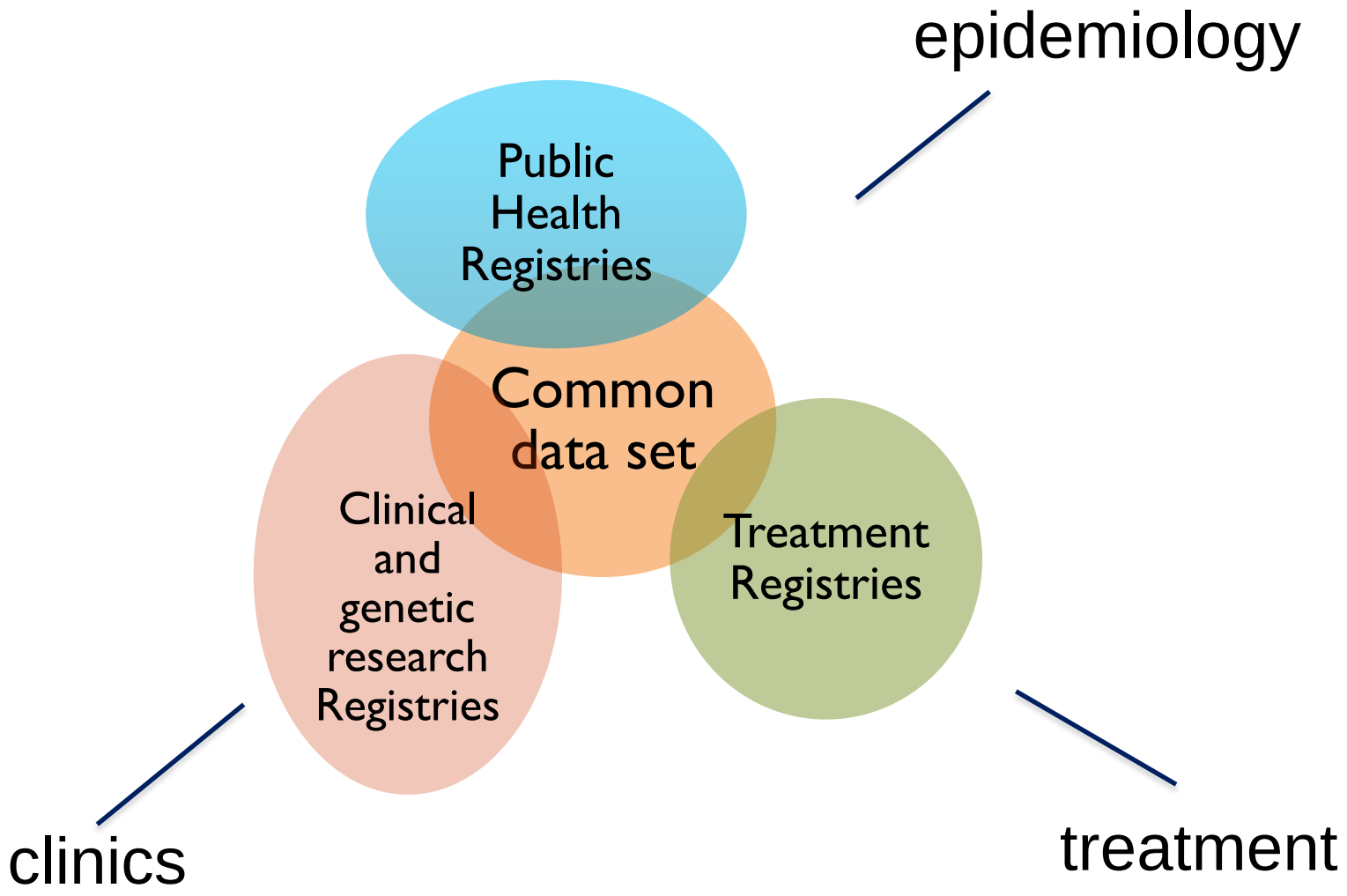
RDs PLATFORM PROTOTYPE



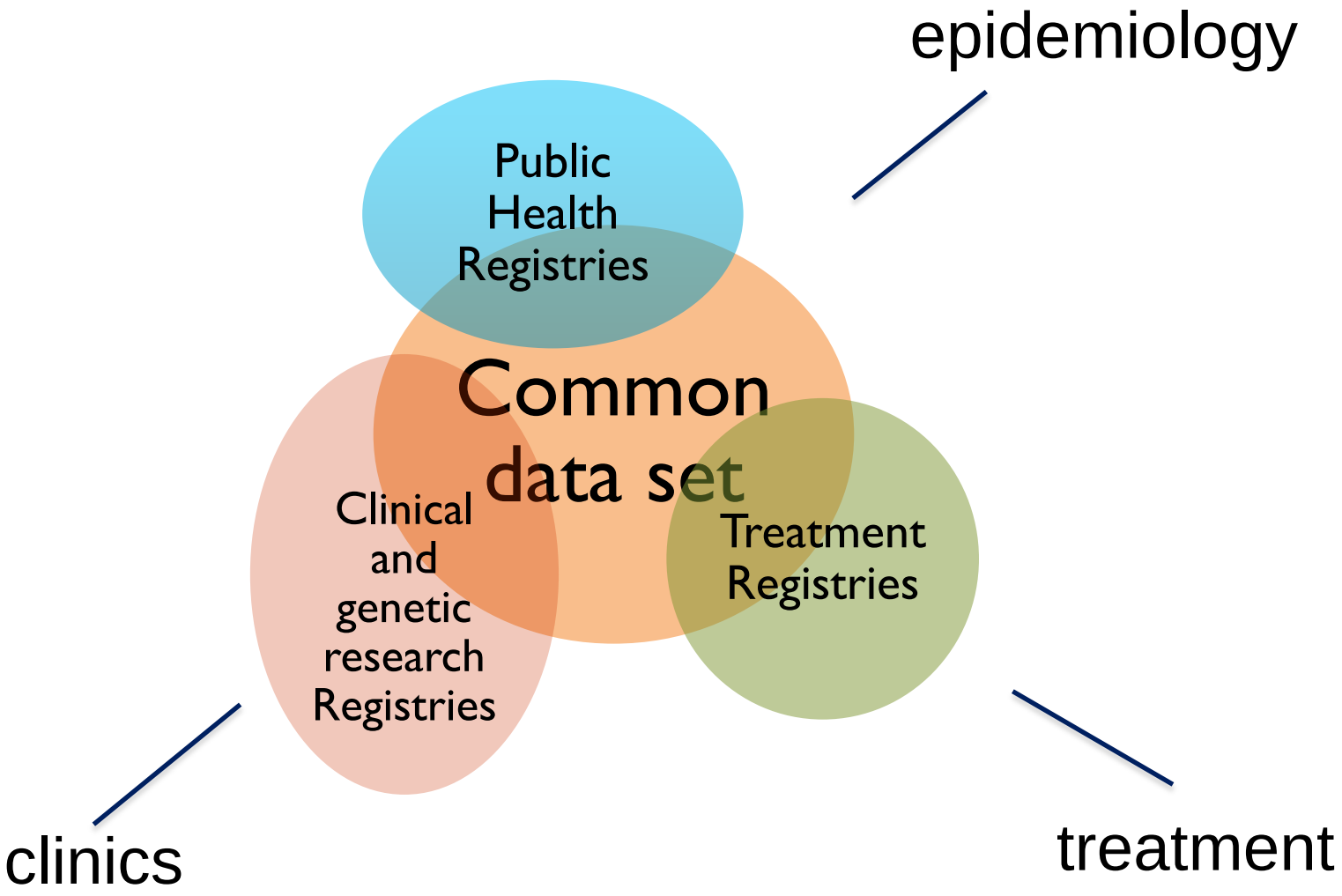
RDs PLATFORM PROTOTYPE



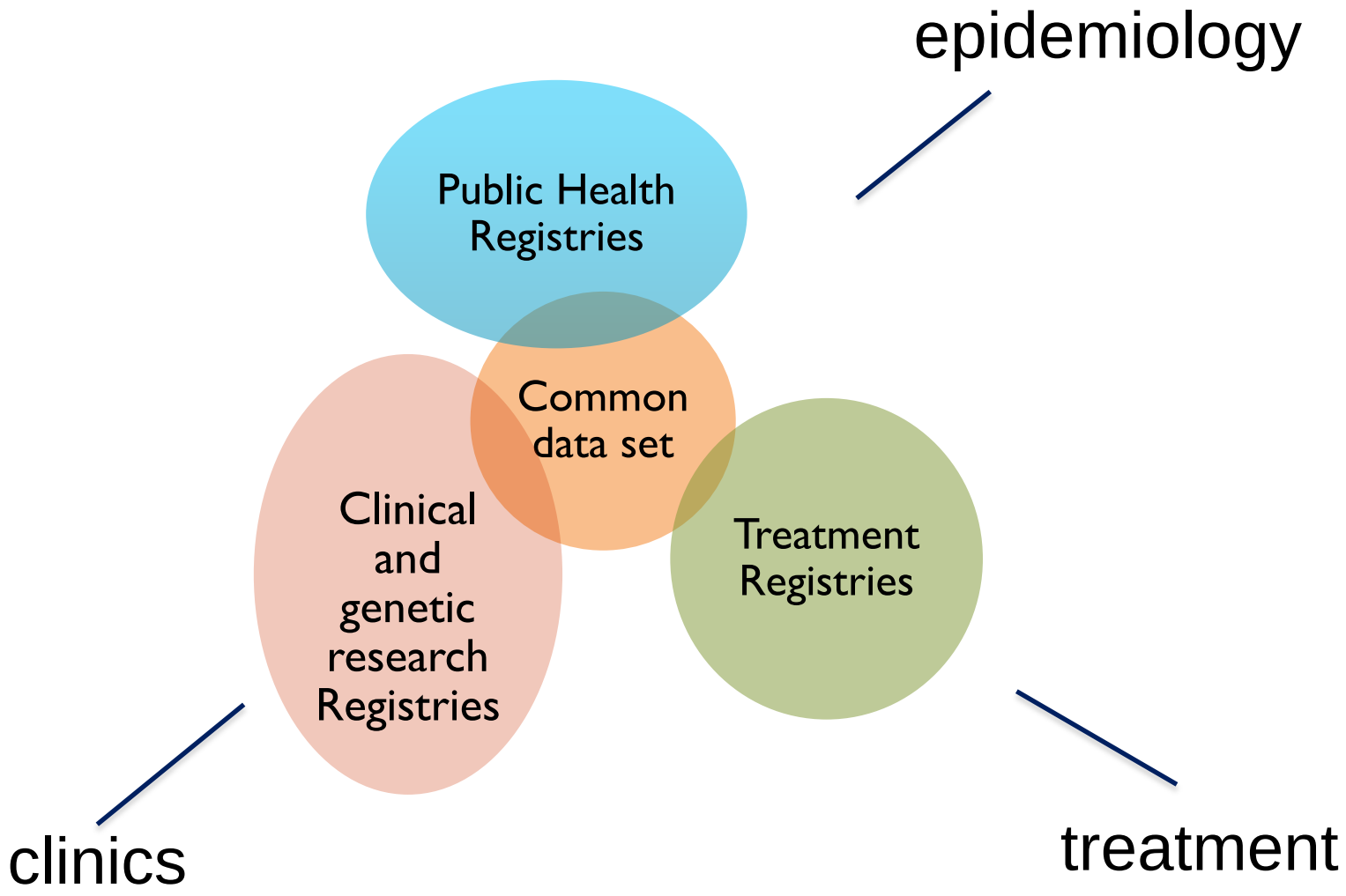
RDs PLATFORM PROTOTYPE



RDs PLATFORM PROTOTYPE



RDs PLATFORM PROTOTYPE



Thank you very much for your attention

on behalf of the EPIRARE staff

fabrizio.bianchi@ifc.cnr.it