



VICT3R

Developing and implementing Virtual Control Groups
to reduce animal use in Toxicology Research



Identifying common study designs for robust Virtual Control Groups generation

Paolo Piraino¹, Stefan Anca² Guillemette Duchateau-Nguyen³

¹ Organon SRL, Bucharest, Romania

² Universitat Oberta de Catalunya, Spain

³ F. Hoffmann-La Roche Ltd, Computational Sciences Center of Excellence, Roche Innovation Center Basel, Switzerland



This project is supported by the Innovative Health Initiative Joint Undertaking (IHI JU) under grant agreement No 101172693. The JU receives support from the European Union's Horizon Europe research and innovation programme and COCIR, EFPIA, Europa Bio, MedTech Europe, and Vaccines Europe and Instem Scientific Limited. Funded by the European Union, the private members, and those contributing partners of the IHI JU. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the aforementioned parties. Neither of the aforementioned parties can be held responsible for them.

The VICT3R Project

Overview



37 Partner organisations from 12 countries

- 6 academic institutions
- 8 Small & medium-sized enterprises (SMEs)
- 23 industry partners (plus 8 associated partners linked to them).
- + collaboration with 4 CROs conducting animal studies

More details about participants available at VICT3R.eu
[VICT3R](https://VICT3R.eu) | [IHI Innovative Health Initiative \(europa.eu\)](https://IHI-Innovative-Health-Initiative.europa.eu)

Project at a Glance

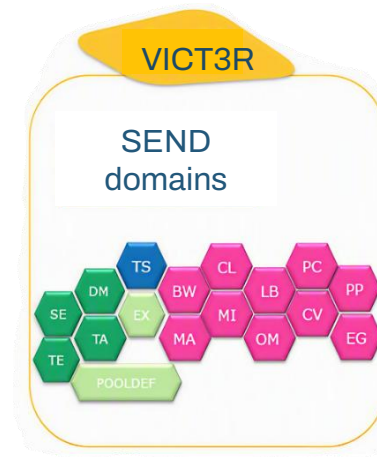
- Kick-off: 1 September 2024
- Finalization: 29 February 2028
- Funding call: HORIZON-JU-IHI-2023-05-01
- Total Budget: $\approx$ €31M
- EU Funding: €13,524,750
- Industry & Contributing partners: €16,639,020

VICT3R Database (V2.4)

~140 k animals

Species	Studies	Animals	Animals (%)	Strains	Male	Female	Missing sex
RAT	5,382	98,189	70.2%	19	50,836	46,980	373
MOUSE	1,046	24,847	17.8%	44	11,897	12,911	39
DOG	1,059	7,582	5.4%	2	3,861	3,670	51
MONKEY	839	6,055	4.3%	5	3,025	3,006	24
RABBIT	271	2,524	1.8%	5	554	1,918	52
MINIPIG	51	362	0.3%	3	197	165	0
GUINEA PIG	16	305	0.2%	3	224	81	0
PIG	11	99	0.1%	1	55	44	0

VICT3R data uses the CDISC SEND format which organizes the nonclinical data into distinct domains



Examples of domains

DM (Demographics): Stores basic subject details like species, strain, sex, age, and treatment arm.

LB (Laboratory Test Results): Captures clinical pathology data such as hematology and clinical chemistry measurements.

MI (Microscopic Findings): Holds histopathology data describing tissue slide observations under a microscope.

From Historical Controls to Virtual Control Group



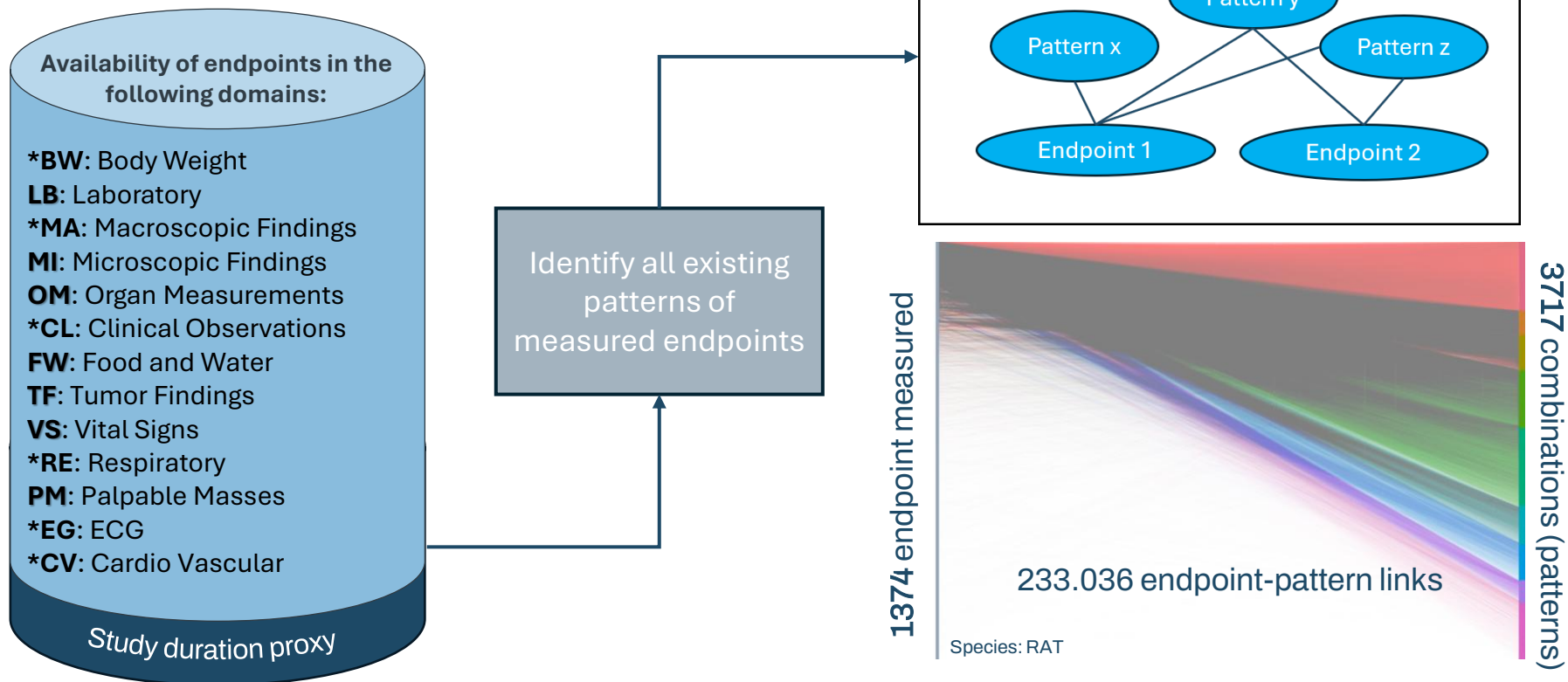
The challenge

- Generating a VCG for a new study requires identifying enough comparable animals* with multivariate endpoint profiles matching the ones required by the study.
- VCGs built of animals being tested for the same patterns of measured endpoints is crucial to preserve the within-animal correlations among findings, to ultimately generate a biologically coherent, and statistically robust data.
- However, even in a large historical-control database, only a small subset of animals may share the required endpoint-measurement pattern, resulting in a limited donor pool concentrated in a few studies — and increased risk of source-related bias.
- We therefore categorized the VICT3R database into study types on the base of the pattern of measured endpoints to identify the use-cases where reliable VCG can be generated.

**(same strain, sex, route of administration, ... and other experimental setting variables)*

Data & Methods

by SPECIES



**domain availability only, single tests, specimen or category not specified.*

Community detection

Find groups containing both studies (patterns) and tests (endpoints) that are more strongly connected to one another than expected.



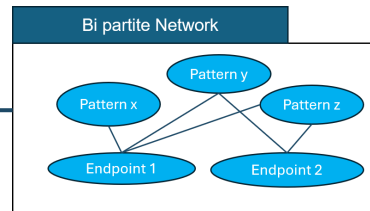
Heatmaps (endpoint profiles vs studies)

	DOG									
	C01	C02	C03	C04	C05	C06	C07	C08	C09	
ICHCT	73.3% 22/30	40.0% 12/30	46.7% 14/30	50.0% 15/30	66.7% 20/30	23.3% 7/30	60.0% 18/30	30.0% 9/30	0.0%	0/30
U.S. EPA1	69.7% 23/33	33.3% 11/33	39.4% 13/33	48.5% 16/33	63.6% 21/33	21.2% 7/33	60.6% 20/33	27.3% 9/33	0.0%	0/33
OECD	66.7% 22/33	27.3% 9/33	27.3% 9/33	42.4% 14/33	60.6% 20/33	12.1% 4/33	57.6% 19/33	15.2% 5/33	0.0%	0/33

Guidelines concordance

Community detection
(Bipartite stochastic block models)

Analysis of communities.

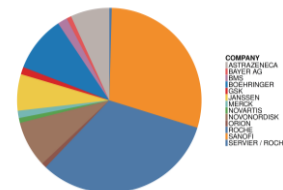


Communities represent
recurring
endpoint-availability
profiles

Frequency tables

Endpoint	Number_of_Studies	Number_of_Studies
M_STOMACH	226	226
M_HEART	226	226
M_SMALL_INTESTINE_ascending	226	226
M_SMALL_INTESTINE_descending	227	226
M_KIDNEY	223	226
M_LIVER	223	224
M_THYROID	221	223
M_TESTES	221	223
M_LARGE_INTESTINE_COLON	218	226
M_SPLEEN	216	217
M_PANCREAS	216	217
M_PANCREAS	210	212
M_SMALL_INTESTINE_jejunum	204	206
M_EYE	201	203
M_LUNG	196	197
M_Ovary	192	194
M_SMALL_INTESTINE_ileum	190	192
M_THYROID	186	192
M_SMALL_INTESTINE	186	190
M_SMALL_INTESTINE	186	188

External* characterization
(by Sponsor, Route,
Housing, Strain ...)



Domain coverage analysis

* metadata not used to find groups 6

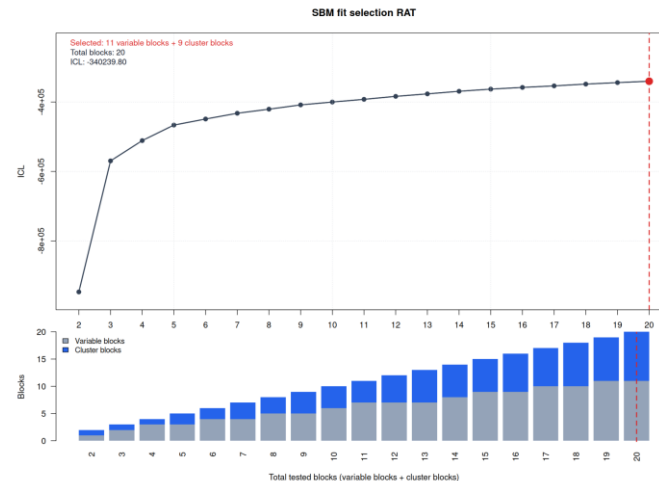
Data-driven study designs



Detected communities (studies sharing a similar pattern of measured endpoints)

Species	N.	Patterns per community M (SD)	Studies per community M (SD)	Ratio
DOG	9	107 (52.6)	117 (56.4)	1,09
MINIPIG	10	4.8 (2.2)	5.1 (2.3)	1,06
MONKEY	8	100.8 (71.2)	104.9 (71.6)	1,04
RAT	9	413 (186.9)	519.4 (514.5)	1,43

N. = number of detected communities, M=mean, SD= Standard Deviation



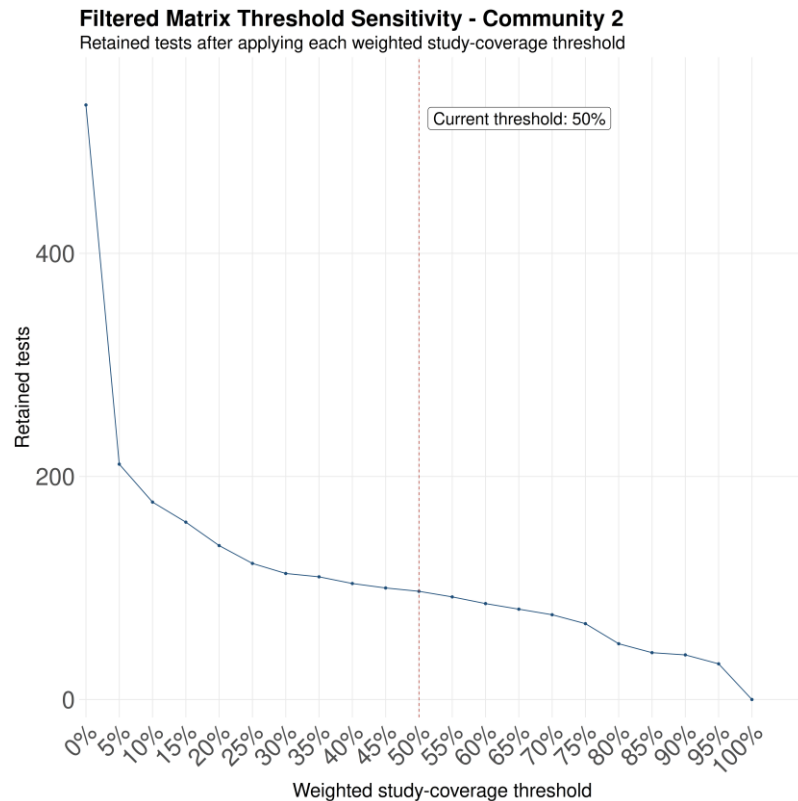
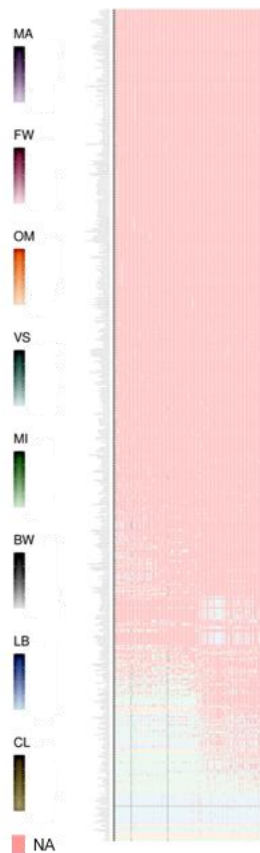
The number of communities is selected automatically by choosing the stochastic block model with the highest ICL (Integrated Completed Likelihood) score, balancing model fit and complexity.

Common patterns vs ad-hoc measurements

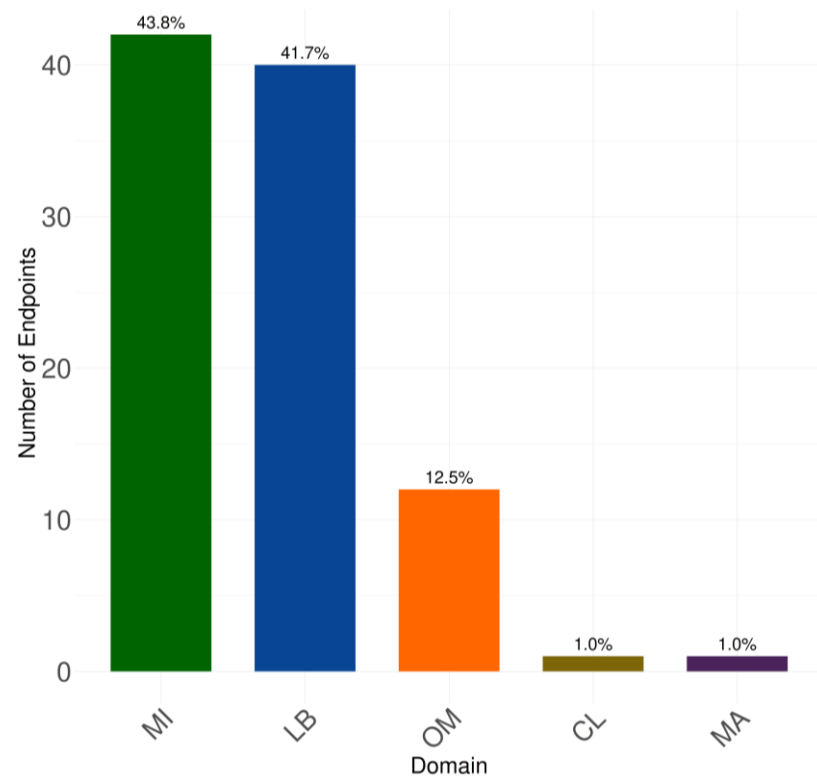
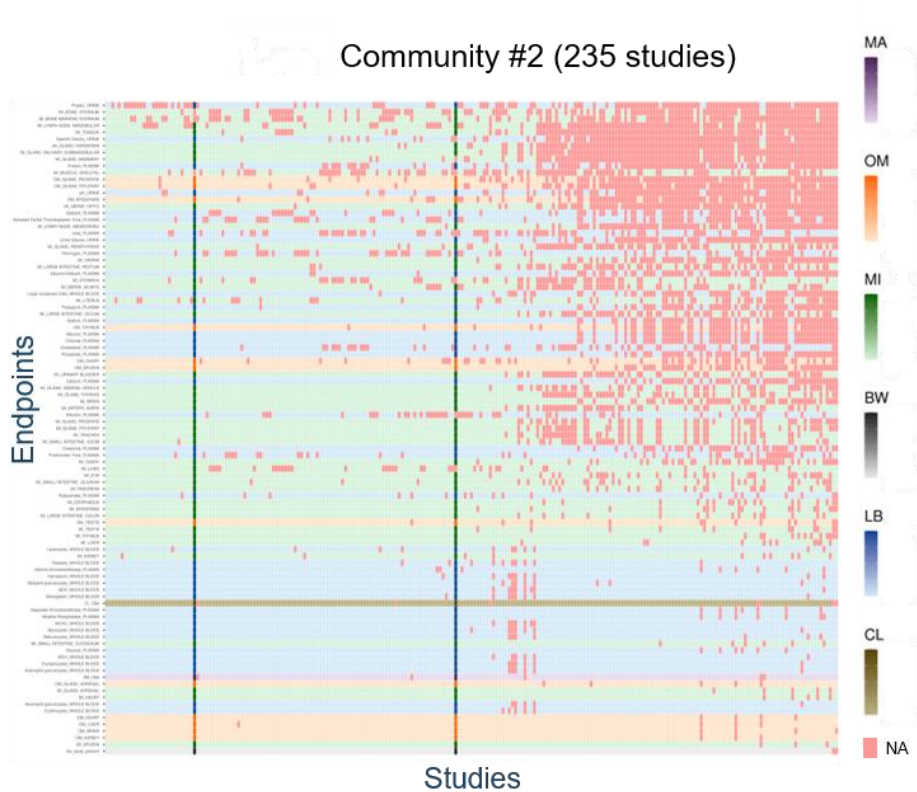
Sensitivity analysis & Filtering

The Vict3r database stores animal findings from ~11.5 hundreds of different tests carried on several specimens... but only a small percent of them are common and co-occur across studies.

We will focus on endpoint-measurements that were observed in at least half of the studies in each community, we omit sparse or rare tests.



Example of a community



Endpoints are selected according a rationale



E.g.: following OECD¹, U.S EPA², or ICH³ guidelines

- International guidelines are internationally harmonized protocols that define how preclinical safety studies should be designed, conducted, monitored, and reported. These guidelines ensure consistency across laboratories and regulatory jurisdictions.
- While guidelines establish a common core, the actual endpoints measured can vary because of:
 - study objective and toxicological question;
 - development phase and regulatory region;
 - expected target-organ or mechanism-related risks;
 - sponsor or laboratory practices;
 - optional, triggered, or scientifically justified additional assessments

In wich extent guidelines are followed? Are observations, and measurements truly consistent across laboratories and regulatory jurisdictions?

Guideline coverage across communities



DOG

IHCPT	73.3% 22/30	40.0% 12/30	46.7% 14/30	50.0% 15/30	66.7% 20/30	23.3% 7/30	60.0% 18/30	30.0% 9/30	0.0% 0/30
U.S. EPA1	69.7% 23/33	33.3% 11/33	39.4% 13/33	48.5% 16/33	63.6% 21/33	21.2% 7/33	60.6% 20/33	27.3% 9/33	0.0% 0/33
OECD	66.7% 22/33	27.3% 9/33	27.3% 9/33	42.4% 14/33	60.6% 20/33	12.1% 4/33	57.6% 19/33	15.2% 5/33	0.0% 0/33
Guideline	C01	C02	C03	C04	C05	C06	C07	C08	C09

MINIPIG

IHCPT	83.3% 25/30	73.3% 22/30	43.3% 13/30	40.0% 12/30	26.7% 8/30	60.0% 18/30	73.3% 22/30	20.0% 6/30	60.0% 18/30	13.3% 4/30
U.S. EPA1	78.8% 26/33	69.7% 23/33	36.4% 12/33	33.3% 11/33	24.2% 8/33	60.6% 20/33	72.7% 24/33	18.2% 6/33	60.6% 20/33	15.2% 5/33
OECD	75.8% 25/33	63.6% 21/33	24.2% 8/33	27.3% 9/33	15.2% 5/33	51.5% 17/33	60.6% 20/33	9.1% 3/33	57.6% 19/33	9.1% 3/33
	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10

MONKEY

IHCPT	43.3% 13/30	73.3% 22/30	30.0% 9/30	60.0% 18/30	23.3% 7/30	43.3% 13/30	63.3% 19/30	0.0% 0/30
U.S. EPA1	36.4% 12/33	69.7% 23/33	24.2% 8/33	60.6% 20/33	21.2% 7/33	36.4% 12/33	63.6% 21/33	0.0% 0/33
OECD	24.2% 8/33	66.7% 22/33	18.2% 6/33	54.5% 18/33	12.1% 4/33	24.2% 8/33	57.6% 19/33	0.0% 0/33
	C01	C02	C03	C04	C05	C06	C07	C08

RAT

IHCPT	73.3% 22/30	43.3% 13/30	26.7% 8/30	60.0% 18/30	56.7% 17/30	0.0% 0/30	23.3% 7/30	26.7% 8/30	0.0% 0/30
U.S. EPA1	69.7% 23/33	36.4% 12/33	21.2% 7/33	60.6% 20/33	57.6% 19/33	0.0% 0/33	21.2% 7/33	30.3% 10/33	0.0% 0/33
OECD	63.6% 21/33	24.2% 8/33	15.2% 5/33	51.5% 17/33	54.5% 18/33	0.0% 0/33	12.1% 4/33	42.4% 14/33	0.0% 0/33
	C01	C02	C03	C04	C05	C06	C07	C08	C09

Community

Domain Coverage

Number of endpoints present in 80% of the studies / total number of endpoints (domain-wise)



DOG

LB - PLASMA	2/3	0/0	18/23	1/3	0/2	0/0	0/2	0/10	0/0
LB - SERUM	16/19	0/0	0/0	11/19	14/19	0/0	15/18	0/1	0/0
LB - URINE	1/3	1/5	1/6	0/0	0/2	0/0	0/0	0/0	0/0
LB - WHOLE BLOOD	14/15	12/15	15/15	14/14	13/21	26/33	13/13	11/14	0/0
MI	35/46	32/46	22/46	36/40	0/0	0/10	0/0	0/0	0/0
OM	6/12	6/13	9/13	6/7	0/11	0/0	0/0	0/0	0/0
CL	0/1	1/1	1/1	0/0	0/1	1/1	0/1	0/0	0/0
MA	1/1	1/1	1/1	1/1	0/1	1/1	0/0	0/0	0/0
BW	1/1	1/1	1/1	1/1	1/1	1/1	0/1	1/1	0/0
	C01 (n=210)	C02 (n=61)	C03 (n=106)	C04 (n=96)	C05 (n=120)	C06 (n=28)	C07 (n=157)	C08 (n=101)	C09 (n=88)

MONKEY

LB - PLASMA	19/22	3/3	0/0	0/3	0/0	18/22	0/2	0/0
LB - SERUM	0/0	18/19	0/0	17/19	0/0	0/0	18/19	0/0
LB - URINE	2/5	3/3	0/2	0/0	0/0	0/4	0/0	0/0
LB - WHOLE BLOOD	15/15	15/16	11/17	14/16	16/30	15/16	13/15	0/0
MI	37/53	38/46	34/48	21/47	0/5	0/11	0/0	0/0
OM	11/13	7/13	8/10	0/0	0/7	0/10	0/3	0/0
CL	1/1	1/1	0/1	1/1	0/1	1/1	0/1	0/0
MA	1/1	1/1	0/1	1/1	0/1	0/1	0/1	0/0
BW	1/1	1/1	1/1	1/1	1/1	1/1	1/1	0/0
	C01 (n=87)	C02 (n=251)	C03 (n=114)	C04 (n=64)	C05 (n=32)	C06 (n=37)	C07 (n=144)	C08 (n=77)

MINIPIG

LB - PLASMA	3/3	3/3	16/23	0/0	0/0	1/3	0/22	0/0	0/1	0/0
LB - SERUM	17/20	19/19	0/0	0/0	0/0	12/18	0/14	0/0	16/20	6/6
LB - URINE	2/8	4/4	2/5	7/8	1/3	0/0	3/10	0/0	0/0	0/0
LB - WHOLE BLOOD	16/17	14/14	15/15	13/13	12/16	10/15	10/40	12/14	14/15	0/0
MI	39/57	55/58	37/52	54/56	21/47	30/55	0/0	0/0	0/0	0/0
OM	8/14	11/12	12/14	12/13	0/0	0/0	9/12	0/0	0/0	0/0
CL	1/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1	0/0
MA	1/1	1/1	1/1	1/1	1/1	1/1	1/1	0/1	0/1	0/0
VS	0/1	0/0	0/0	0/0	0/1	0/0	0/1	0/1	0/0	0/0
FW	0/0	0/1	0/0	0/0	0/0	0/1	0/0	0/0	0/0	0/0
BW	1/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1	0/0
RE	0/0	0/0	0/0	0/0	0/0	0/0	0/1	0/0	0/0	0/0
EG	0/5	4/6	0/5	0/0	0/0	0/0	0/8	0/0	0/0	0/0
	C01 (n=8)	C02 (n=6)	C03 (n=6)	C04 (n=2)	C05 (n=2)	C06 (n=6)	C07 (n=2)	C08 (n=7)	C09 (n=5)	C10 (n=4)

RAT

LB - PLASMA	0/3	7/21	0/0	0/3	0/0	0/0	0/5	0/0	0/0
LB - SERUM	16/18	0/0	0/0	17/23	16/18	0/0	0/0	13/19	0/0
LB - URINE	0/3	0/4	0/1	0/0	0/0	0/0	0/0	0/0	0/0
LB - WHOLE BLOOD	15/16	14/15	15/15	15/16	13/15	0/0	0/14	0/0	0/0
MI	36/49	20/42	6/27	7/19	0/0	17/41	0/0	0/0	0/0
OM	8/12	6/12	0/10	0/6	0/0	0/9	0/1	0/0	0/0
CL	0/1	1/1	0/0	0/0	0/1	0/0	0/1	0/1	0/1
MA	0/1	1/1	0/0	0/0	0/1	0/1	0/0	0/0	0/0
BW	1/1	1/1	1/1	1/1	1/1	0/1	1/1	0/1	1/1
	C01 (n=615)	C02 (n=233)	C03 (n=292)	C04 (n=491)	C05 (n=731)	C06 (n=317)	C07 (n=339)	C08 (n=174)	C09 (n=525)

Link between study duration & endpoint profiles

The study duration proxy was included as node in the network (link is probability)

DOG									
Single-day/acute (<=1 day)				0.8%		2.2%	2.8%	6.8%	
Short (2-7 days)	0.5%	1.6%	0.9%	5.2%	4.8%	3.6%	13.1%	4.6%	16.5%
2 w (8-16 days)	10.4%	13.1%	19.8%	36.5%	31.7%	67.9%	31.7%	23.1%	31.6%
4 w (17-35 days)	34.4%	39.3%	40.6%	32.3%	15.9%	25.0%	24.0%	29.6%	19.5%
6-8 w (36-70 days)	17.0%	6.6%	12.3%	2.1%	23.8%		8.2%	7.4%	3.0%
13 w (71-120 days)	18.9%	21.3%	15.1%	6.2%	8.7%		5.5%	14.8%	3.8%
6 m (121-210 days)	2.4%	4.9%	1.9%	4.2%	6.3%		1.1%	3.7%	0.8%
9 m (211-320 days)	12.3%	8.2%	8.5%		7.1%	3.6%	0.5%	7.4%	0.8%
12 m (321-400 days)	4.2%	4.9%	0.9%	6.2%	0.8%			4.6%	0.8%
Longer than 12 m (>400 days) (missing study duration)				1.0%			0.5%		0.8%
				6.2%			13.1%	1.9%	15.8%
	C01 (n=210)	C02 (n=61)	C03 (n=106)	C04 (n=96)	C05 (n=120)	C06 (n=28)	C07 (n=157)	C08 (n=101)	C09 (n=88)

MONKEY								
Single-day/acute (<=1 day)		0.4%	2.6%	14.1%		5.4%	6.0%	3.9%
Short (2-7 days)		1.6%		3.1%	3.1%	2.7%	4.7%	9.8%
2 w (8-16 days)	3.4%	6.7%	10.4%	40.6%	34.4%	32.4%	39.6%	20.6%
4 w (17-35 days)	35.6%	37.9%	29.6%	17.2%	25.0%	32.4%	18.1%	35.3%
6-8 w (36-70 days)	23.0%	6.7%	11.3%	15.6%	15.6%	16.2%	6.0%	6.9%
13 w (71-120 days)	14.9%	23.7%	25.2%	7.8%	12.5%	5.4%	15.4%	6.9%
6 m (121-210 days)	11.5%	9.1%	11.3%		9.4%		3.4%	2.9%
9 m (211-320 days)	10.3%	13.4%	9.6%	1.6%			2.0%	1.0%
12 m (321-400 days)	1.1%	0.4%				2.7%	0.7%	
Longer than 12 m (>400 days) (missing study duration)							0.7%	
						2.7%	3.4%	12.7%
	C01 (n=87)	C02 (n=251)	C03 (n=114)	C04 (n=64)	C05 (n=32)	C06 (n=37)	C07 (n=144)	C08 (n=77)

MINIPIG										
Single-day/acute (<=1 day)									20.0%	
Short (2-7 days)	12.5%									14.3%
2 w (8-16 days)			16.7%		50.0%	66.7%		14.3%	20.0%	28.6%
4 w (17-35 days)	25.0%	66.7%	16.7%	50.0%	50.0%		50.0%	42.9%	60.0%	28.6%
6-8 w (36-70 days)	25.0%	16.7%				33.3%	50.0%	42.9%		28.6%
13 w (71-120 days)	12.5%		66.7%	50.0%						
9 m (211-320 days)	25.0%	16.7%								
	C01 (n=8)	C02 (n=6)	C03 (n=6)	C04 (n=2)	C05 (n=2)	C06 (n=6)	C07 (n=2)	C08 (n=7)	C09 (n=5)	C10 (n=4)

RAT									
Single-day/acute (<=1 day)	0.5%			0.2%	2.7%		5.9%	8.0%	6.8%
Short (2-7 days)	4.3%	2.6%	15.0%	18.6%	27.7%	4.2%	13.3%	26.9%	10.2%
2 w (8-16 days)	21.5%	21.3%	18.4%	61.6%	30.7%	32.4%	26.4%	30.8%	30.4%
4 w (17-35 days)	29.4%	33.2%	14.1%	14.9%	17.9%	21.9%	24.1%	19.4%	27.6%
6-8 w (36-70 days)	12.8%	8.5%	4.6%	1.8%	9.9%	11.4%	5.1%	1.5%	7.7%
13 w (71-120 days)	15.2%	17.0%	6.1%	1.2%	5.8%	7.2%	8.2%	5.0%	3.9%
6 m (121-210 days)	13.0%	10.2%	4.0%	1.2%	2.9%	2.1%	4.6%	1.5%	1.5%
9 m (211-320 days)	2.7%	3.4%	1.5%		0.6%	8.1%	0.8%		0.6%
12 m (321-400 days)	0.2%	1.3%			0.5%		1.8%		0.3%
Longer than 12 m (>400 days) (missing study duration)	0.5%	1.7%	0.3%		0.1%	2.4%	3.6%		0.5%
		0.9%	35.9%	0.6%	1.3%	10.2%	6.2%	7.0%	10.5%
	C01 (n=615)	C02 (n=233)	C03 (n=292)	C04 (n=491)	C05 (n=731)	C06 (n=317)	C07 (n=339)	C08 (n=174)	C09 (n=525)

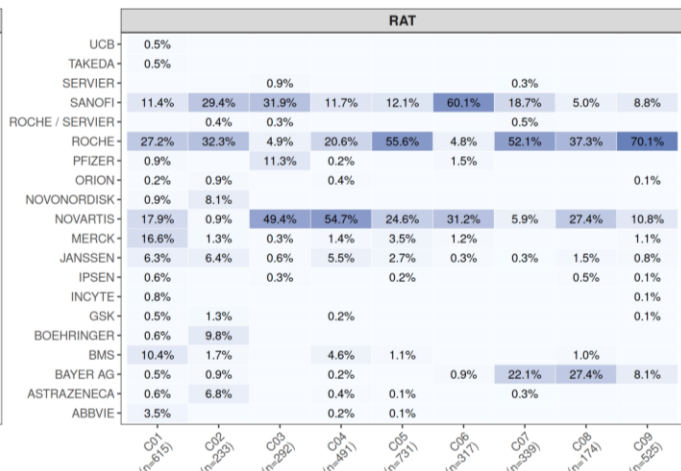
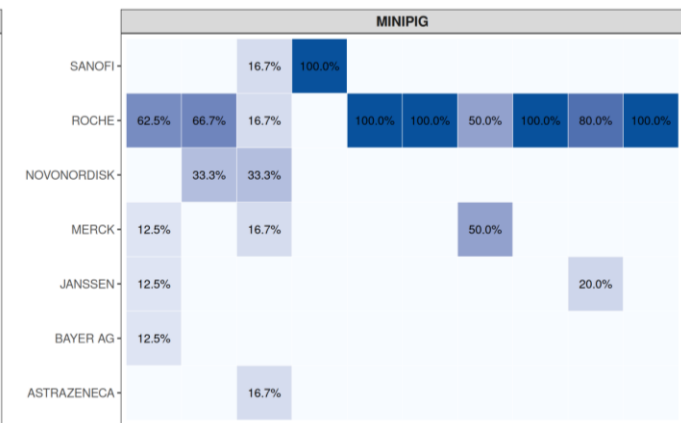
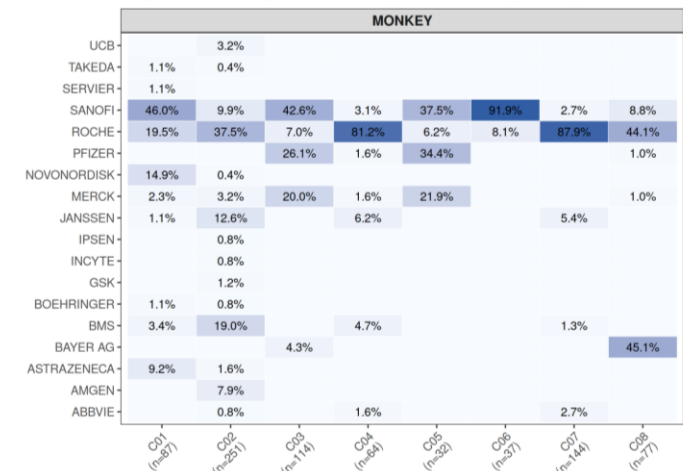
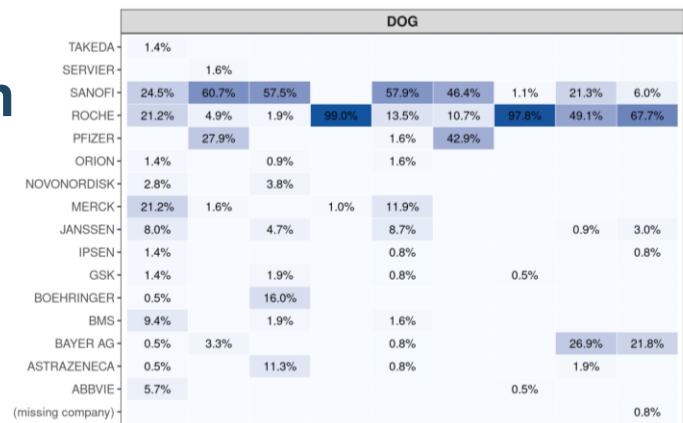
Community

External characterization

Sponsor was not included as node in the network (descriptive analysis)

Sponsor is a proxy for several study setting parameters that may influence endpoint variability.

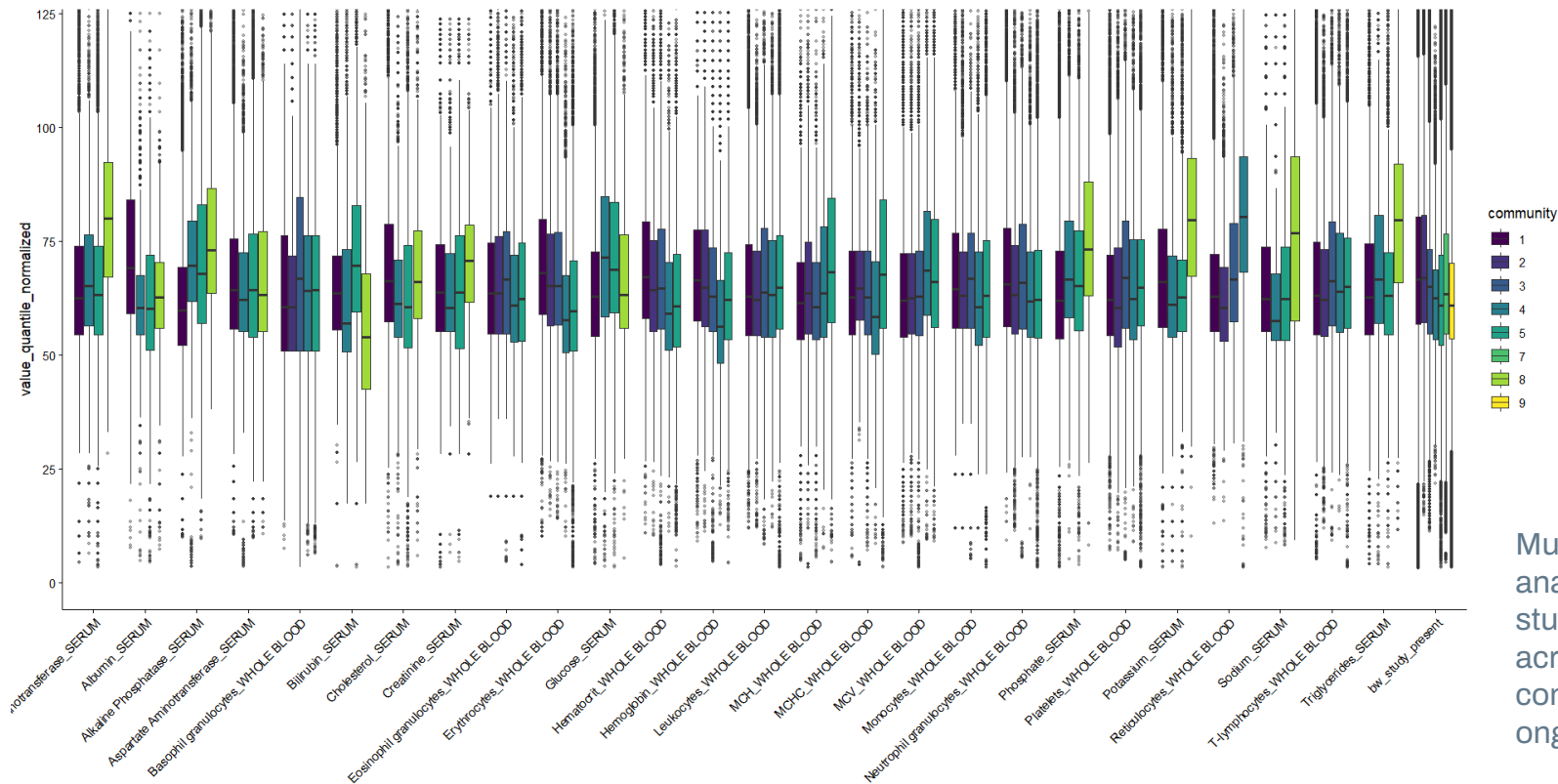
Sponsor representation is uneven across endpoint communities, with some dominated by one or two sponsors.



Same endpoint in different communities



Is the distribution of measured values the same across communities?



Multivariate analysis of within-study correlations across communities is ongoing

Endpoint set (community) 1 - RAT



Comprehensive regulatory-style repeat-dose general-toxicity studies (n=632)

LB—whole blood: eosinophils, reticulocytes, monocytes, T-lymphocytes, erythrocytes, leukocytes, platelets, hemoglobin, neutrophils, MCH, MCHC, MCV, large unstained cells, basophils, hematocrit.

LB—serum: bilirubin, creatinine, albumin/globulin ratio, chloride, sodium, globulin, AST, ALP, albumin, ALT, potassium, glucose, phosphate, triglycerides, calcium, cholesterol.

MI: eye, liver, lung, mandibular lymph node, prostate gland, stomach, spleen, kidney, thymus, pituitary, urinary bladder, uterus, ovary, epididymis, thyroid, heart, pancreas, tongue, mesenteric lymph node, aorta, brain, esophagus, adrenal, mammary, parathyroid, seminal vesicle, cecum, colon, optic nerve, sciatic nerve, duodenum, ileum, jejunum, testis, trachea, vagina.

OM: brain, testis, adrenal, kidney, liver, spleen, thymus, heart.

BW: present.

Broad clinical pathology, organ weights and full histopathology. Strong coherence: ALT/AST/ALP/bilirubin with liver MI and liver weight; creatinine/electrolytes with kidney MI and weight; hematology with spleen/thymus/lymph-node MI and spleen/thymus weights; lipids/glucose with liver and pancreas MI.

Endpoint set (community) 4 - RAT



Core-organ repeat-dose general-toxicity studies (n=505)

LB—whole blood: eosinophils, reticulocytes, monocytes, T-lymphocytes, erythrocytes, leukocytes, platelets, hemoglobin, neutrophils, MCH, MCHC, MCV, large unstained cells, basophils, hematocrit.

LB—serum: bilirubin, creatinine, albumin/globulin ratio, chloride, sodium, globulin, AST, ALP, albumin, ALT, potassium, total protein, glucose, phosphate, triglycerides, calcium, cholesterol.

MI: liver, lung, spleen, kidney, thymus, heart, adrenal.

BW: present.

Hematology/clinical chemistry panel comparable to 1 but with a narrower histopathology and organ measurements panel. Strong liver assessments (ALT, AST, ALP, bilirubin, proteins/lipids plus liver MI), kidney (creatinine/electrolytes plus kidney MI), and hematology (blood counts plus spleen/thymus MI). OM is common but less than the threshold to be listed (77.23%)

Endpoint set (community) 2 - RAT



Highly standardized repeat-dose general-toxicity studies (n=235)

LB—whole blood: eosinophils, reticulocytes, monocytes, T-lymphocytes, erythrocytes, leukocytes, platelets, hemoglobin, neutrophils, MCH, MCHC, MCV, basophils, hematocrit.

LB—plasma: prothrombin time, ALT, triglycerides, creatinine, glucose, ALP, AST.

MI: eye, liver, lung, prostate gland, spleen, kidney, thymus, pituitary, ovary, epididymis, heart, pancreas, esophagus, adrenal, colon, duodenum, ileum, jejunum, testis, trachea.

OM: brain, testis, adrenal, kidney, liver, heart.

CL: observations.

MA: observations.

BW: present.

Probably following a coherent laboratory/CRO template using plasma rather than serum chemistry. CL and MA are almost universal alongside LB, MI, OM and BW. Liver and kidney chemistry align with corresponding MI and organ weights; hematology aligns with spleen/thymus MI; prothrombin time adds hepatic/coagulation context.

Endpoint set (community) 3 - RAT



Hematology-centred repeat-dose or abbreviated general-toxicity phenotype (n=326)

LB—whole blood : eosinophils, reticulocytes, monocytes, T-lymphocytes, erythrocytes, leukocytes, platelets, hemoglobin, neutrophils, MCH, MCHC, MCV, large unstained cells, basophils, hematocrit.

MI: liver, lung, spleen, kidney, heart, pancreas.

BW: present

Body weight and a restricted core microscopic organ set. Frequent blood-cell findings and spleen MI assessment. Liver, kidney and pancreas are examined microscopically, but not always with chemistry analytes, so biochemical–morphological corroboration for those organs is weak. OM reaches 80.37% as a domain, yet no individual organ measurement reaches the 80% core threshold, indicating heterogeneous organ-weight panels.

Endpoint set (community) 5-6 - RAT



Inconsistent laboratory / pathology confirmation

5 (n=851)	6 (n=333)
LB—whole blood: eosinophils, monocytes, T-lymphocytes, erythrocytes, leukocytes, platelets, hemoglobin, neutrophils, MCH, MCHC, MCV, basophils, hematocrit. LB—serum: bilirubin, creatinine, albumin/globulin ratio, chloride, sodium, AST, ALP, albumin, ALT, potassium, total protein, glucose, phosphate, triglycerides, calcium, cholesterol. BW: present. No core MI, OM, CL, MA, FW or other tests.	MI: liver, lung, spleen, kidney, thymus, uterus, ovary, epididymis, heart, pancreas, brain, esophagus, adrenal, colon, duodenum, ileum, jejunum, testis. No other individual core tests.
Clinical-pathology centered with incomplete pathology delivery. The laboratory panel for liver, kidney, metabolism and hematology. However, MI is available in only 48.41% and OM in 53.35%. Reticulocytes and globulin, core in some comprehensive communities, are not core here.	Pathology-centred or pathology-only/legacy data extracts. Major organ covered, including gastrointestinal tract and reproductive organs; LB is present in only 3.30%, so liver, kidney and hematology lesions usually cannot be correlated with clinical chemistry or hematology. BW (72.37%) is less represented. This could reflect a special pathology study.

Endpoint set (community) 7-9 - RAT



Heterogeneous or sparse endpoint sets

7 (n=390)	8 (n=201)	9 (n=1850)
BW: present. LB is universal as a domain, but no individual LB test reaches 80%.	LB—serum: bilirubin, creatinine, sodium, AST, ALP, albumin, ALT, potassium, total protein, glucose, phosphate, triglycerides, cholesterol. No core hematology, pathology or BW test.	BW: present. No other core test or observation.
Heterogeneous laboratory/clinical-observation studies, possibly mixed short-duration, range-finding or differently standardized general-toxicity designs. LB in every study but no shared core analyte indicates incompatible laboratory panels. CL is relatively common (73.85%), whereas MI and MA are sparse.	Serum-clinical-chemistry-centred studies, potentially abbreviated toxicity/tolerability studies supporting liver, renal/electrolyte and metabolic surveillance. Biochemical changes generally lack consistent morphological or organ-weight corroboration (MI 32.34%, OM 44.28%).	Heterogeneous mixture of studies in which body weight and clinical observations (61.62%) were the most consistently captured endpoints. LB, MI and OM each occur in fewer than 18%. This could include short exploratory, range-finding, special-purpose studies, or incomplete/legacy extracts.

Conclusions and outlook



VCG feasibility depends on the joint endpoint profile and donor diversity—
not simply on the overall size of the database

- Recurring, biologically coherent study designs based on the endpoints measured together within animals were identified using graph theory.
- The VICT3R database provides substantial historical-control coverage for several common toxicology endpoint profiles, establishing candidate use cases for VCG generation and validation.
- Sparse, heterogeneous, and source-concentrated profiles were also identified. For these profiles, hybrid designs (Concurrent Control and VCG) may be more appropriate.
- Communities (studies-endpoints groups) are being used in ongoing multivariate analyses to quantifying the within-animal relationships among endpoints and inform imputation methods that preserve cross-endpoint correlation, with the aim of producing biologically coherent VCGs across a broader range of endpoints.
- The present work also identifies where future data contributions by sponsors should be focused to enrich currently underrepresented profiles and progressively expand the range of feasible VCG applications.



Thanks!

Contact us:

info@vict3r.eu



vict3r.eu



Co-funded by
the European Union

This project is supported by the Innovative Health Initiative Joint Undertaking (IHI JU) under grant agreement No 101172693.

The JU receives support from the European Union's Horizon Europe research and innovation programme and COCIR, EFPIA, Europa Bio, MedTech Europe, and Vaccines Europe and Instem Scientific Limited.

Funded by the European Union, the private members, and those contributing partners of the IHI JU. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the aforementioned parties. Neither of the aforementioned parties can be held responsible for them.

