



UNIVERSITY OF TARTU
Institute of Computer Science

Estonian personalized medicine initiative: polygenic risk scores, pharmacogenetics and rare mutations

Jaak Vilo

vilo@ut.ee

@ESTbiobank @jaakvilo

■ cs.ut.ee

🐦 @UniTartuCS

f ut.ics

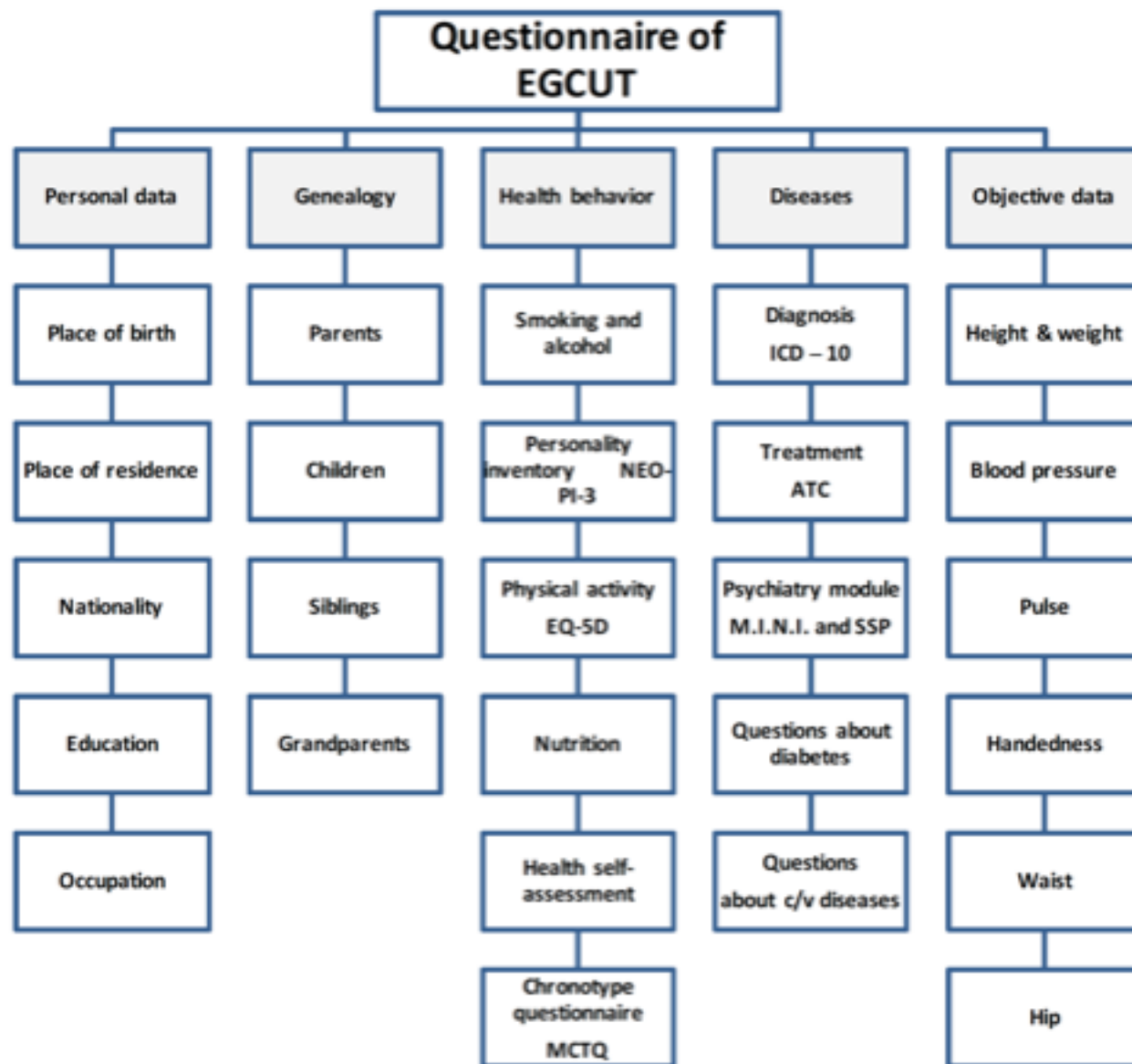
Estonian Biobank (est. 2000): prospective, longitudinal, volunteer-based biobank



- **52,000 participants + 100,000 during 2018**
- **Health records, diet, physical activity, etc.**
- **DNA, plasma and cell samples**

- **Estonian Human Genes Research Act**
- **Broad informed consent**
- **Open for research: Clear access rules**





Qure Browser – a web based data entry user interface

Firefox - Qure Browser - Tuberculosis Register

[1] Persons > [1] JOHN SCHMIDT (33609010275) 000007 > [A] Case 2012-05-22 primary (00000701) > [A.3] Treatment cards > [B] Treatment card 2012-05-22 primary > [B.1] Data card info (0000070101)

Persons Case 2012-05-22 primary (00000701) Death Register data Search of cases Treatment cards search

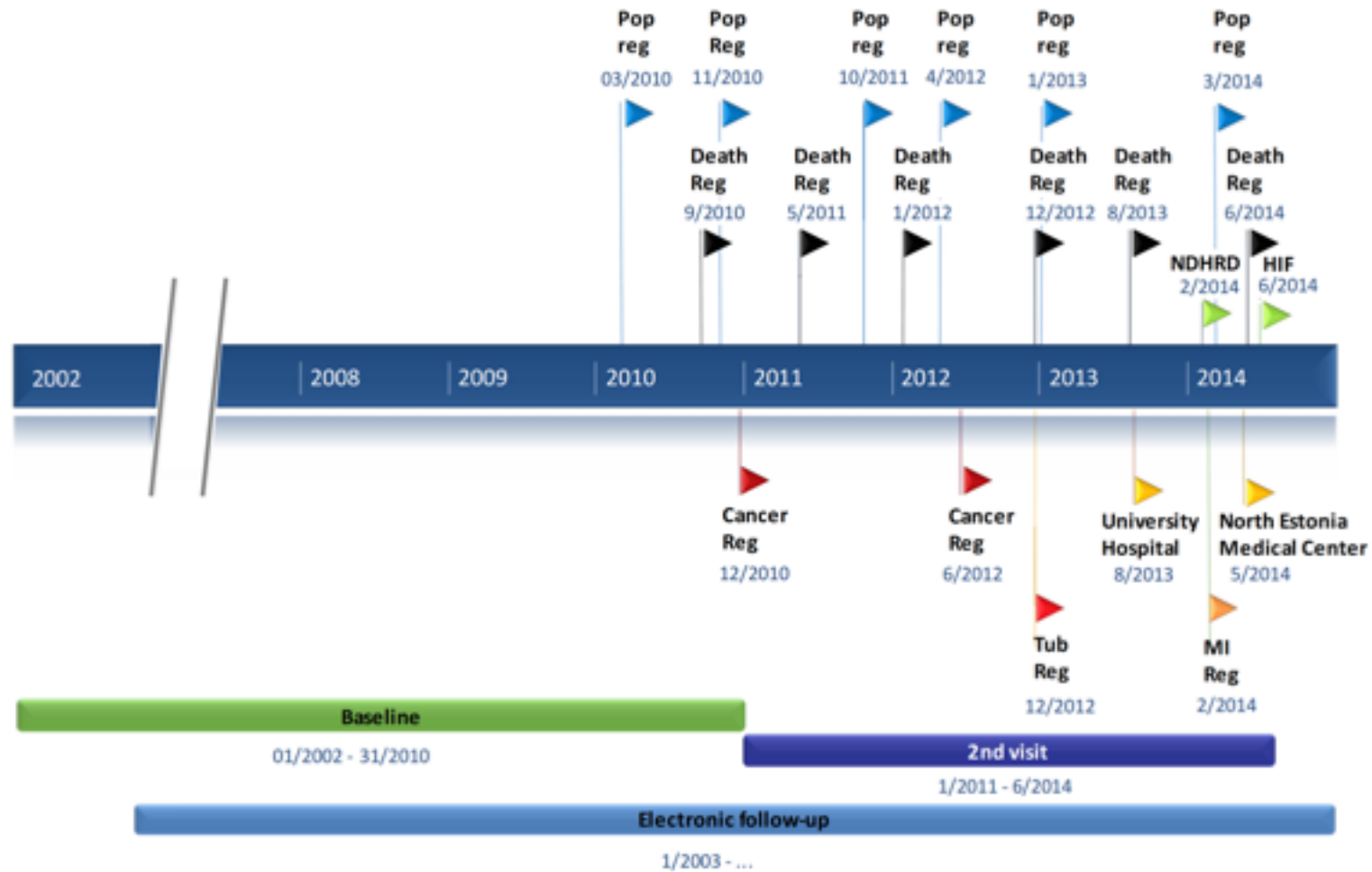
Basic data Treatment cards Councils Comorbidities Surgery Medications Drugs adverse events Lab notifications

Data card info (0000070101)

[B.1] DATA CARD INFO (0000070101)

B.1.1	Baseline	22.05.2012	B.1.2	End of treatment	22.05.2012	B.1.3	Date of diagnosis	dd.mm.yyyy	B.1.4	Days of treatment	
B.1.5*	Baseline condition	primary	B.1.6	Final Condition	treatment ended	B.1.7	Place of treatment				
B.1.8	Start of outpatient therapy	dd.mm.yyyy	B.1.9	End of outpatient therapy	dd.mm.yyyy	B.1.10	Start of hospital treatment	dd.mm.yyyy	B.1.11	End of hospital treatment	dd.mm.yyyy
B.1.12	Diagnosis of tuberculosis	A15.2 Select... [A15.2] Tuberculosis pulmonum histologic confirmata. Condiciones sub A15.0 datae, histologic	B.1.13	Other diagnosis	Select... (Unselected)						
B.1.14	Diagnosis of tuberculosis 2	Select... (Unselected)	B.1.15	Other diagnosis 2	Select... (Unselected)						
B.1.16	Diagnosis of death	Select... (Unselected)	B.1.17	Date of death	dd.mm.yyyy						
B.1.18*	Definition of diagnosis	pulmonary tuberculosis	B.1.19	Form	infiltrative	B.1.20	Location				
B.1.21	Sputum smear	0 not done	B.1.22	Sputum culture	1 +	B.1.23	Destruction	no	B.1.24	Quantiferon test	
B.1.25	HIV		B.1.26	Date of HIV test	dd.mm.yyyy	B.1.27	MDR at baseline	no	B.1.28	MDR ravi lõpul	
B.1.29	BK findings in other material?		B.1.31	Histology							
B.1.33*	County date	22.05.2012	B.1.34*	County	Tallinn	B.1.35*	County medical	Other Doctor (2)			
B.1.36	Medical institution		B.1.37	TOR							
B.1.40	Filler		B.1.41	Filler institution		B.1.42	Filling date	dd.mm.yyyy			
B.1.43	Cause of treatment failure		B.1.44	Basis of		B.1.45	Age	75			
B.1.46*	Show statistics	1 yes	B.1.47	Remarks							

Timeline: linking to registries



Diagnoses in database – 380,000

ICD-10 code, diagnosis	Classes of reliability					Number of diagnoses in the database
	1	2	3	4	Unknown ^a	
C50 Malignant neoplasm of breast	110	37	4	82	49	282
C53 Malignant neoplasm of cervix	37	16	2	63	33	151
C61 Malignant neoplasm of prostate	102	25	2	40	20	189
E05 Thyrotoxicosis (hyperthyroidism)	329	57	14	281	100	781
E10 Insulin-dependent diabetes mellitus	222	75	5	108	48	458
E11 Non-insulin-dependent diabetes mellitus	1432	209	15	369	213	2238
E66 Obesity	987	270	152	538	38	1985
E78 Disorders of lipoprotein metabolism and other lipidaemias	2013	157	19	385	42	2616
F20 Schizophrenia	238	34	2	33	4	311
F32 Depressive episode	1176	276	82	1844	413	3791
F33 Recurrent depressive disorder	410	149	24	224	99	906
G20 Parkinson's disease	87	21	3	21	1	133
G40 Epilepsy	270	76	13	184	42	585
I10 Essential (primary) hypertension	3748	554	161	2771	164	7398
I11 Hypertensive heart disease	4321	710	39	640	36	5746
I21 Acute myocardial infarction	644	57	5	185	48	939
J45 Asthma	834	218	29	796	109	1986
K25 Gastric ulcer	487	80	9	784	182	1542
K26 Duodenal ulcer	861	122	31	1096	413	2523
L40 Psoriasis	403	133	34	599	42	1211
M80 Osteoporosis with pathological fracture	102	35	3	57	4	201
M81 Osteoporosis without pathological fracture	304	76	21	559	20	980
N18 Chronic renal failure	199	36	0	12	3	250
N20 Calculus of kidney and ureter	400	64	14	407	187	1072
Total	19 716	3 487	6 83	12 078	2 310	38 274

Cornerstones of national Health-IT

1. ID code and electronic identity

2. X-Road infrastructure to link information systems

- Logging, time-stamping by block-chain security
- Central reusable information service and registries

3. Health data for research and new IT services

- Nation-wide **billing information** (all procedures, costs)
- **Biobank** – general open informed consent form
- **Electronic health records** – medical notes, etc.
Hospital and GP EHR-s submit to and use the central database
- **E-Prescription** >99% of prescriptions **digital only**
- Nation-wide **health registries**

2002



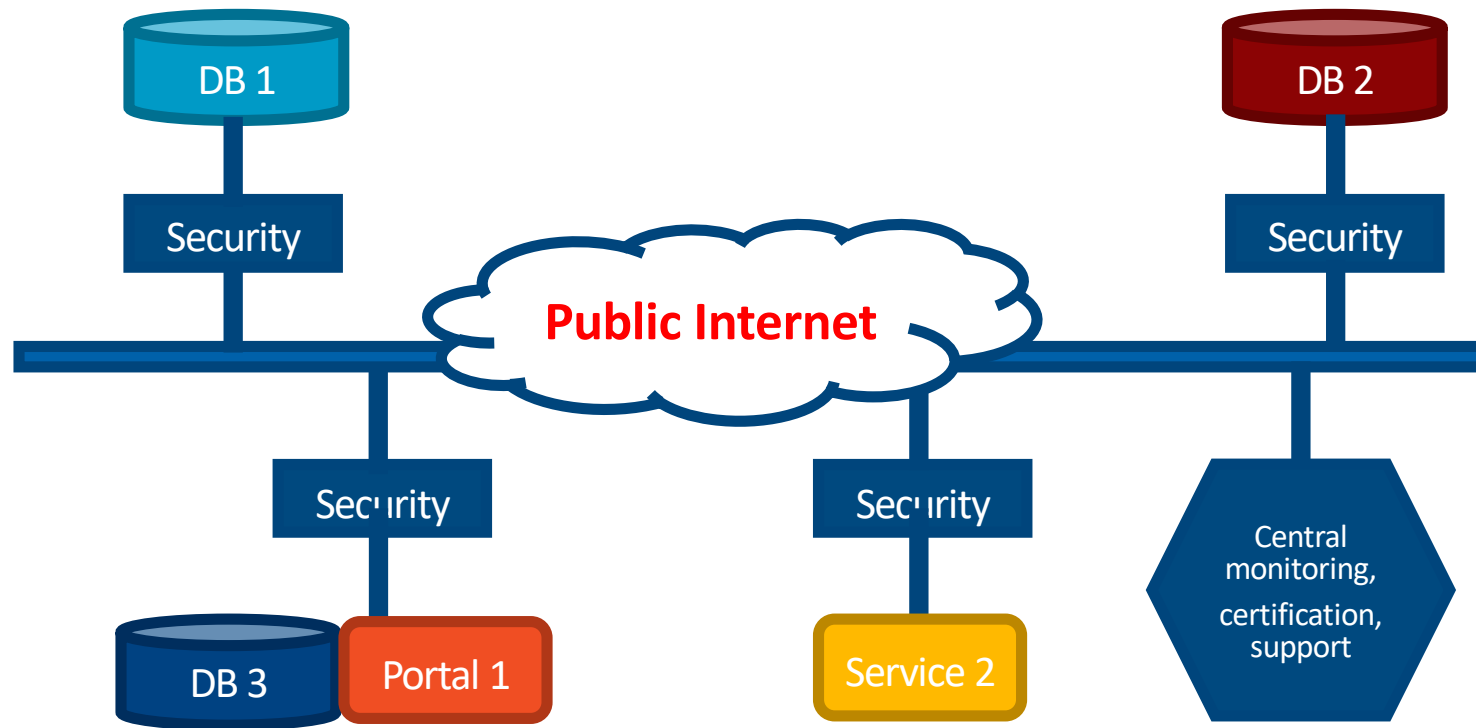
2007



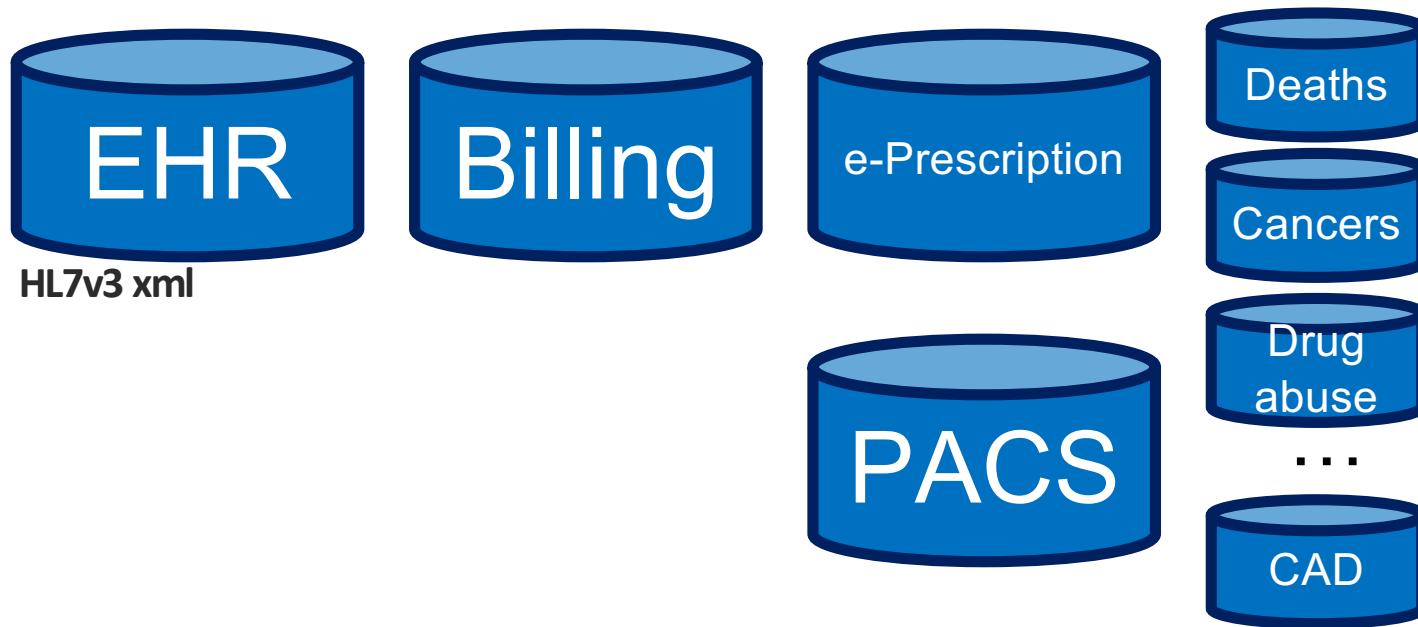
2017



Linked databases - X-Road common bus



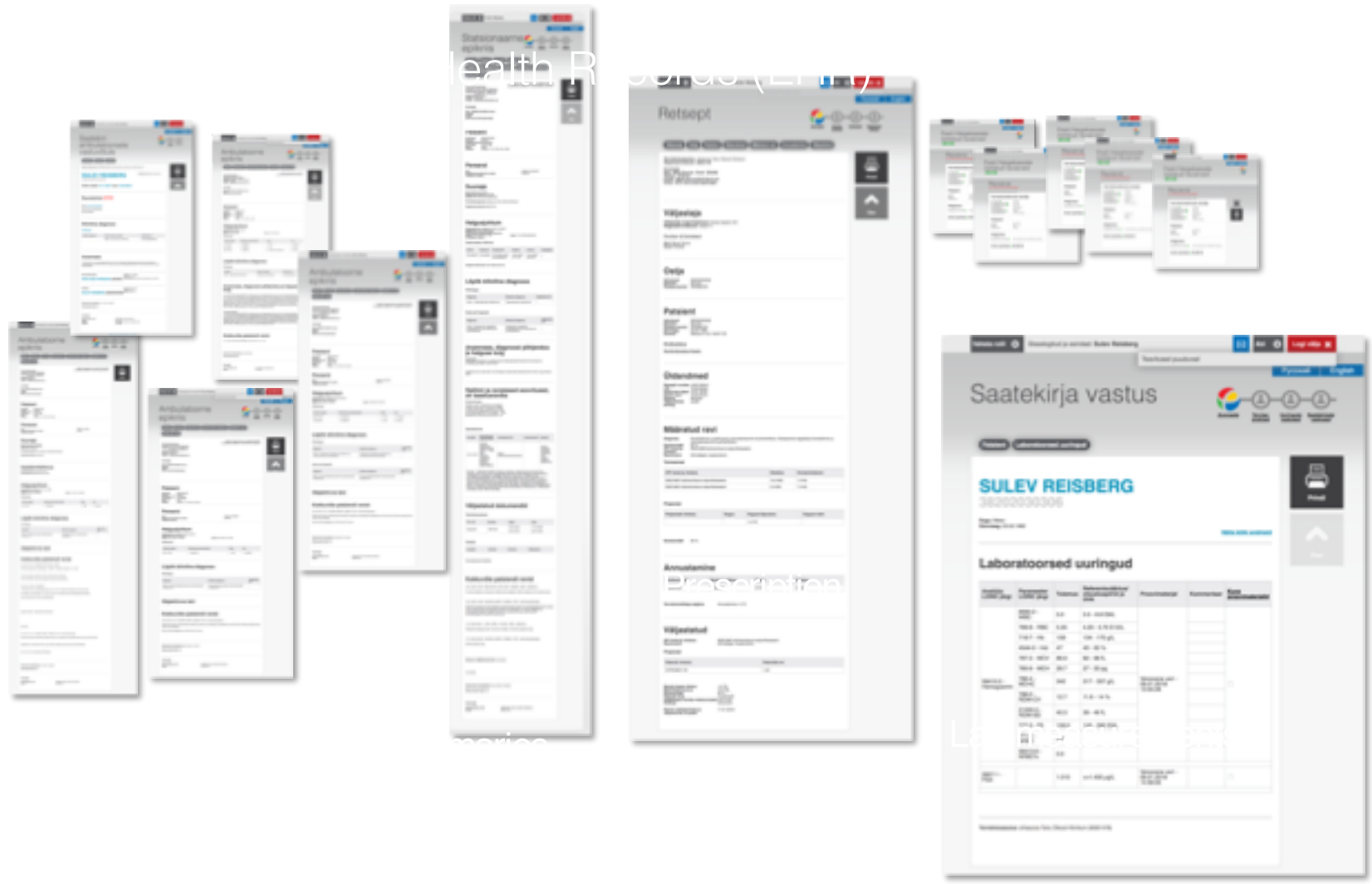
Real-life health data – secondary use



Nr of documents in E-Health DB 1.53 M individuals (06.03.2017)

Document type	Nr. of documents in E-health
Outpatient notes	15,323,163
Inpatient discharge summaries	1,643,296
Development assessment notices	34,308
Immunization side effects	6
Immunization notes	490,624
Growth notes	130,448
Ambulance Cards	340,274
Home nurse reports	5,634
Examination reports	122,945
Advisory notices	105,224
Pointers to “pictures” in PACS	2,866,152
Referrals	991,790
Referral responses	6,902,871

Health Records (EHR)



Automated information extraction

How PSA measurements have actually been written in medical records:

PSA 03042012 - 0,83ng/ml perearsti poolt .

PSA 2010. 3ng/ml, PSA 2012. 1,53ng/ml . - Bx va

PSA 20105,99 ja 26.01.2012 uuesti .

PSA 2011 oli 0 , 4 nG7ml .

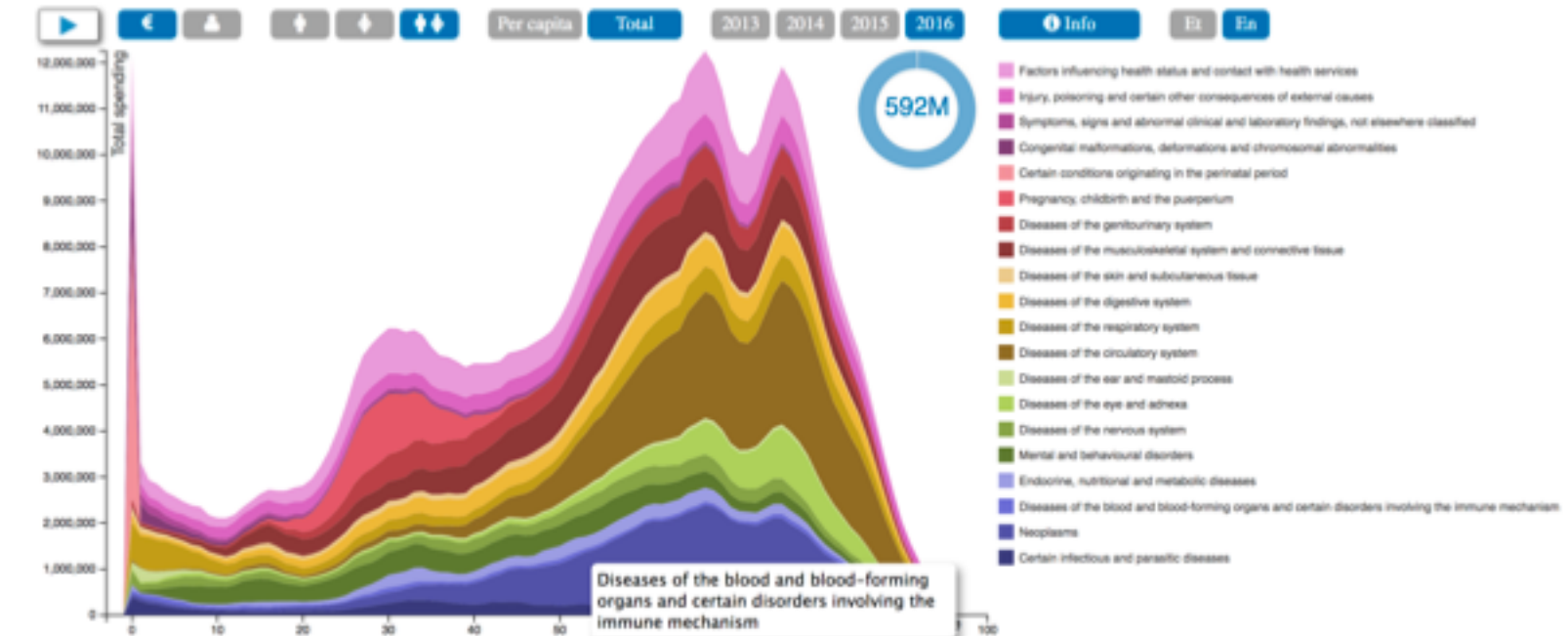
PSA 201222,25ng/ml

PSA 2 aastajooksuldünaamikata , eriuuring

Health data analysis for >1M individuals

- **Text Mining** (Estonian NLP)
- **Machine learning** – predicting context, semantic meaning, semantic similarities,
- **Information extraction** from unstructured data
- Disease **co-morbidities** and **trajectories**
- **Quality indicators**
- **Observational data vs registries (e.g. infarctions)**
- IMI – EMIF-Platform: **EGCUT data and OMOP**
- **Linking Genomics data to EHR** and the needed information architectures

Overview of the Estonian specialized medical care spending and patient counts by age and diagnosis



Source: Estonian Health Insurance Fund

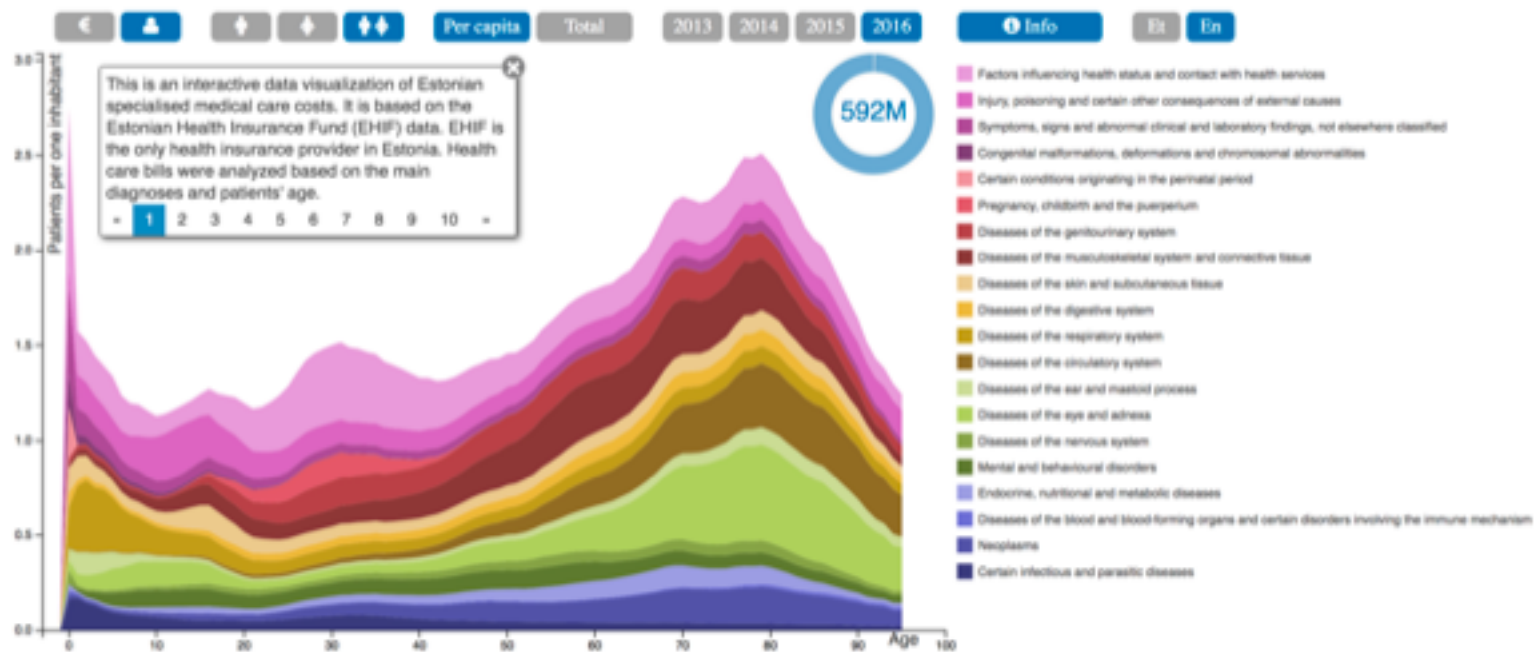


STACC Software Technology and Applications Competence Center



<https://www.stacc.ee/ehif-stacked-area/?lng=En>

Overview of the Estonian specialized medical care spending and patient counts by age and diagnosis



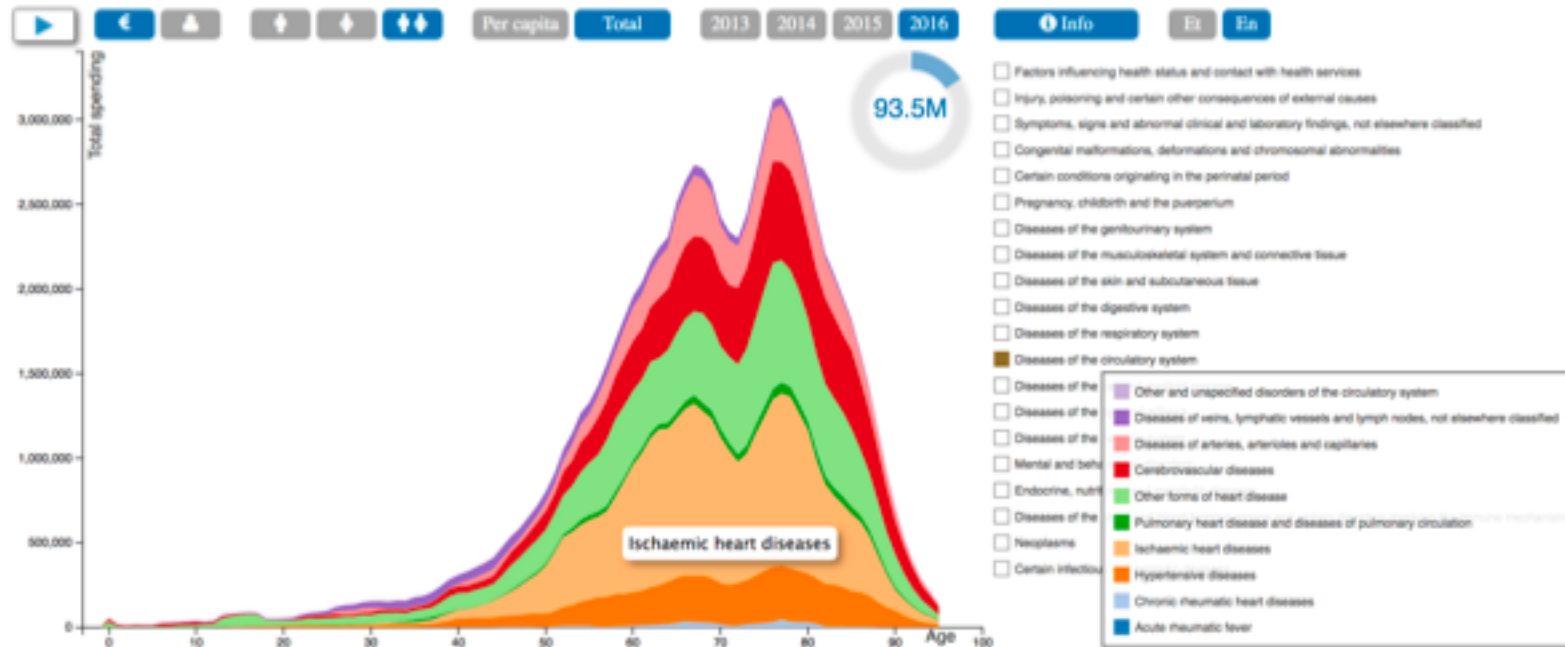
Source: Estonian Health Insurance Fund



STACC Software Technology and Applications Competence Center



Overview of the Estonian specialized medical care spending and patient counts by age and diagnosis



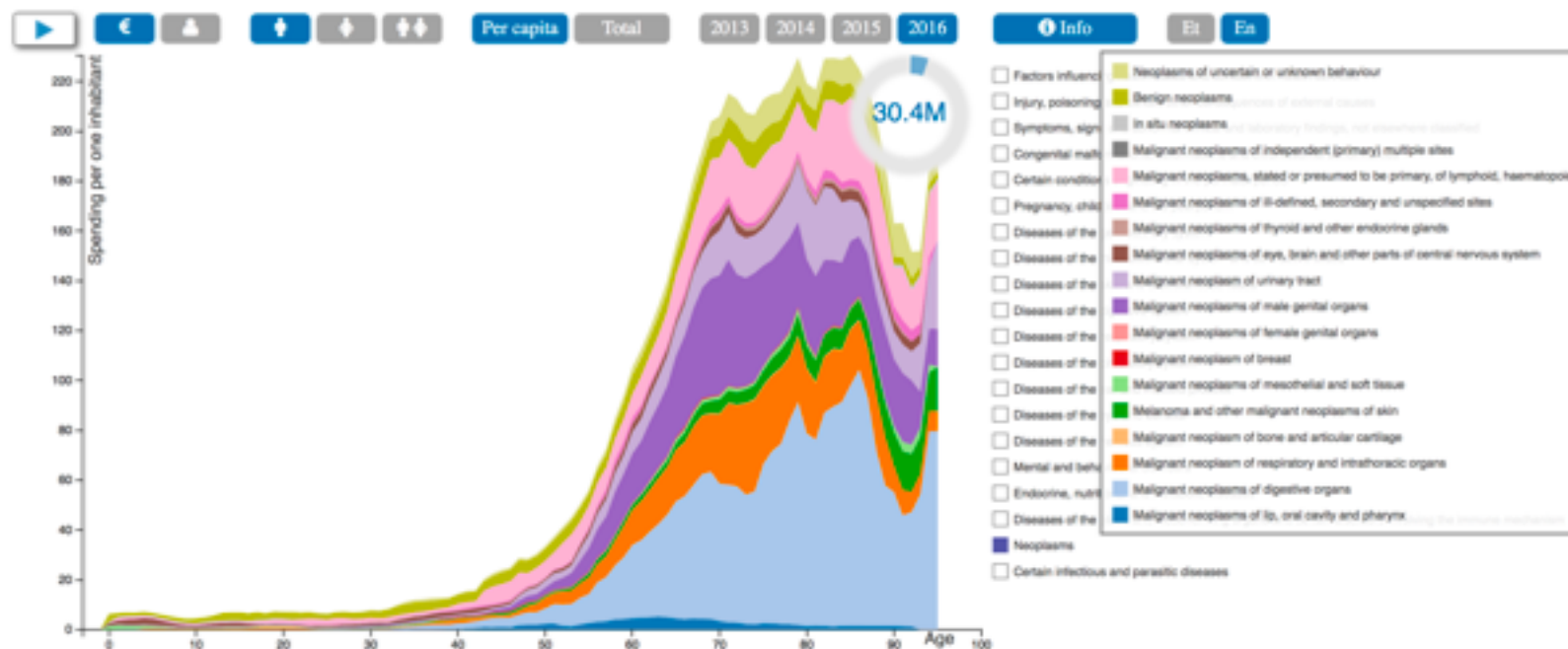
Source: Estonian Health Insurance Fund



STACC Software Technology and Applications Competence Center



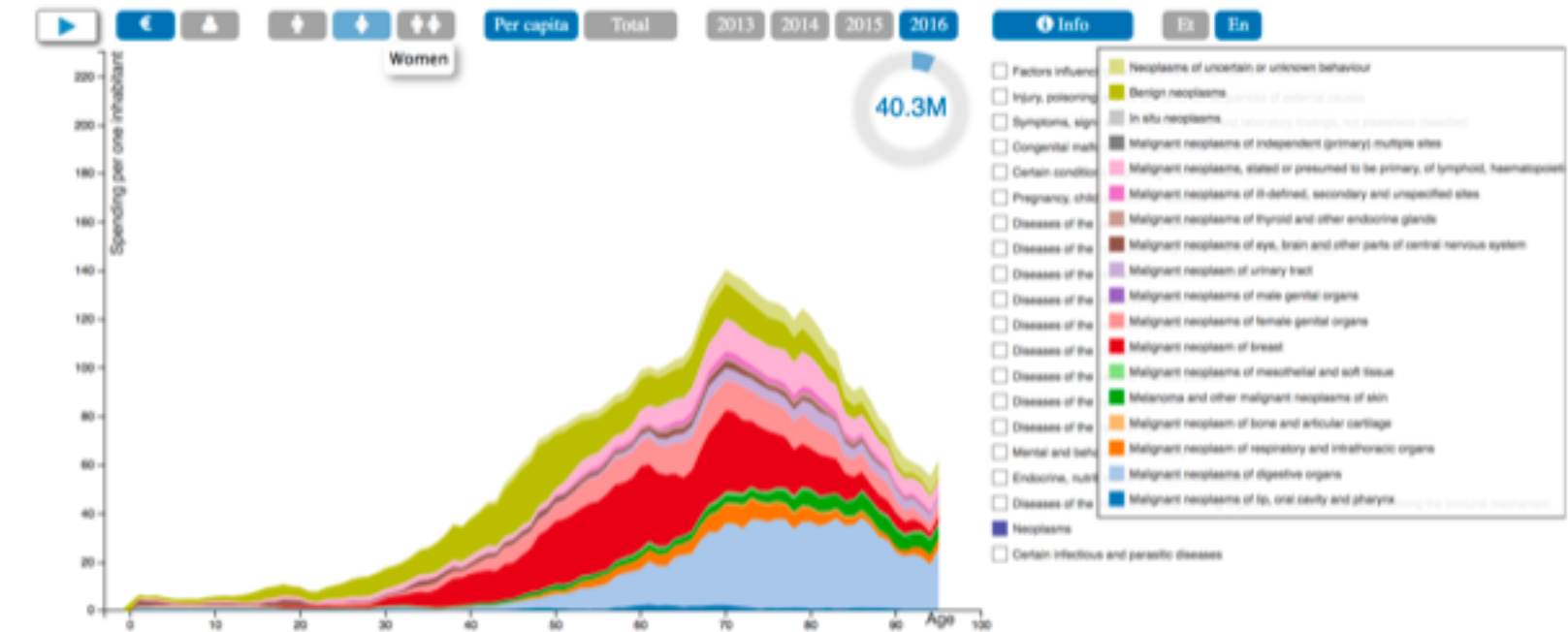
Overview of the Estonian specialized medical care spending and patient counts by age and diagnosis



Source: Estonian Health Insurance Fund



Overview of the Estonian specialized medical care spending and patient counts by age and diagnosis

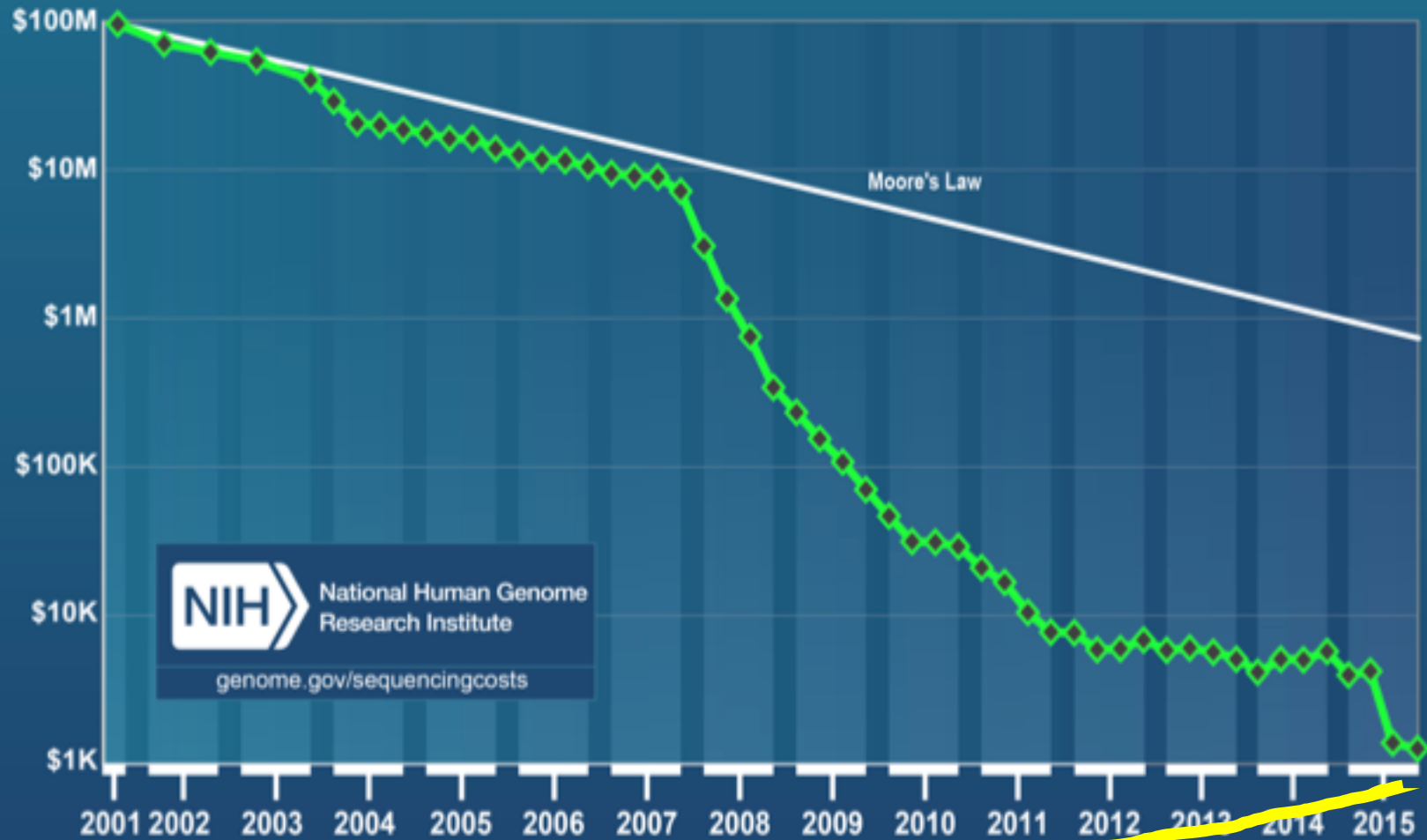


Source: Estonian Health Insurance Fund



<https://www.stacc.ee/ehif-stacked-area/?lng=En>

BENEFIT ~~Cost~~ per Genome





Cold
Spring
Harbor
Laboratory

bioRxiv

THE PREPRINT SERVER FOR BIOLOGY

HOME | ABOUT

New Results

Translating genotype data of 44,000 biobank participants into clinical pharmacogenetic recommendations: challenges and solutions

 Sulev Reisberg, Kristi Krebs, Mart Kals, Reedik Magi, Kristjan Metsalu, Volker M Lauschke, Jaak Vilo, Lili Milani

doi: <https://doi.org/10.1101/356204>

This article is a preprint and has not been peer-reviewed [what does this mean?].

Abstract

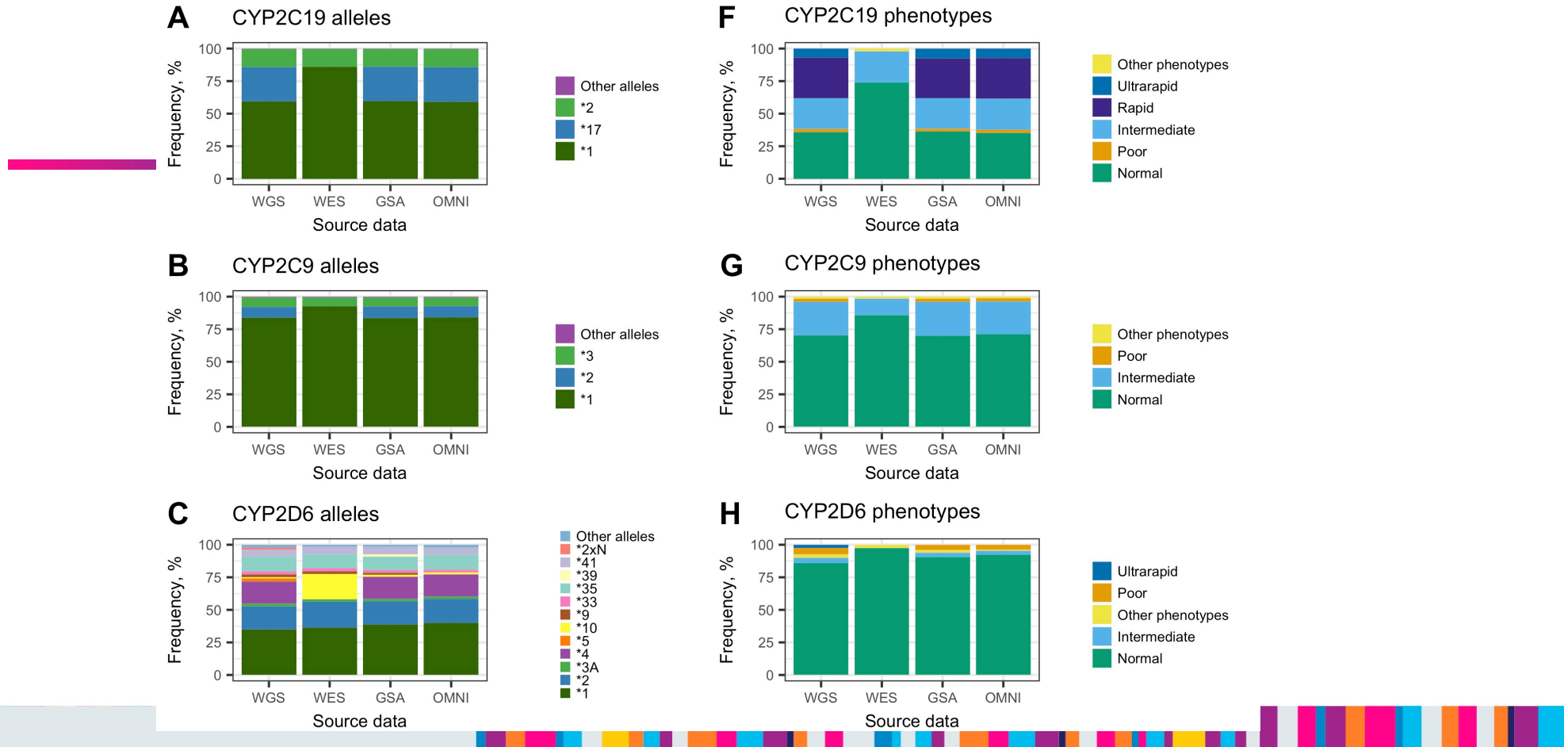
Info/History

Metrics

Supplementary material

 Preview PDF

Pharmacogenetics: calling star alleles & corresponding phenotypes



Pharmacogenetics report + case 2

- 33-y female with depression
- CYP2C19 slow metabolizer,
- 50% dose reduction recommended
- Sertralin and Escitalopram formerly prescribed
- Both discontinued due to side effects: **agitation, aggressiveness, pharyngitis**, etc.

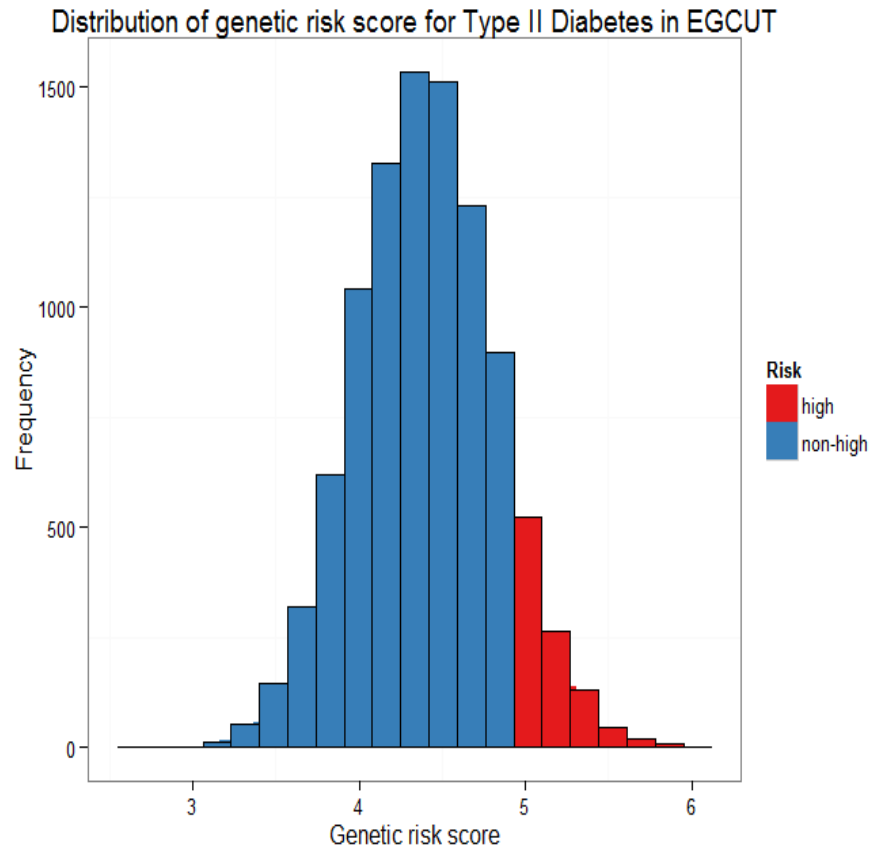
Geen	Genotüüp	Hinnang	Soovitus	Mõjutatud ravimi toimeained
CYP2C19	*2/*2	Aeglane ravimi lagundamine	!	Estsitalopraam, Tsitalopraam, Klopidoogrel, Sertraliin, Vorikonasool, Esomeprasool, Lansoprasool, Pantoprasool, Omeprasool, Klomipramiin, Amitriptiin
CYP2C9	*1/*1	Tavapärane ravimi lagundamine	+	Fenütoiin
CYP2C9; VKORC1	*1/*1; rs9923231 (AA)	Tavapärast madalam doosisoovitus	!	Varfariin
CYP3A5	*3/*3	Aeglane ravimi lagundamine, Tavapärane muster	+	Takroliimus
DPYD	*1/*5	Tavapärane ravimi lagundamine	+	Kapetsitabiin, Fluorouratsiil
IFNL3	rs12979860 (CC)	Tavapärane ravimi toime	+	Alfa-2b-peginterferoon, ribaviin
SLCO1B1	rs4149056 (TT)	Tavapärane müopaatia risk	+	Simvastatiin
TPMT	*1S/*1	Tavapärane ravimi lagundamine	+	Tioguaaniin, Merkaptopuriin, Asatiopriin

+ - Tarvitada tavapärast.
 ! - Tarvitada ettevaatusega, võib vajada doosi muutmist.
 ! - Tarvitada väga ettevaatlikult, oht kõrvaltoimeteks.



	DDD/1000 inhabitants per day (stats from 2016) all of the affected drugs summarised					
Gene	Estonia	Finland	Sweden	Denmark	Norway	Nr of Estonian biobank participants with a prescription
<i>CYP2C19</i>	17.62	43.08	66.83	61.5	38.75	12254
<i>CYP2C9</i>	7.08	16.26	10.7	7.9	7.23	2599
<i>CYP2D6</i>	9.16	15.92	14.29	9.3	15.37	6907
<i>CYP3A5</i>	0.14	0.24	0.5	0	0.47	26
<i>CYP4F2</i>	7.03	16.04	10.23	7.9	7.02	2588
<i>DPYD</i>	0	0	0	0	0	155
<i>IFNL3</i>	0.23	0.04	0.05	0	0	215
<i>SLCO1B1</i>	6.13	40.21	39.19	62.9	39.07	3900
<i>TPMT</i>	0.32	1.41	1.12	0.8	0.87	251
<i>UGT1A1</i>	0	0.03	0.04	0	0.09	215
<i>VKORC1</i>	7.03	16.04	10.23	7.9	7.02	2588
TOTAL	54.74	149.27	153.184	158.2	115.886	37222

Distribution of T2D polygenic risk score



$$PRS = \sum_{i=1}^n w_i \cdot X_i$$

Figure 1

From

Personalized risk prediction for type 2 diabetes: the potential of genetic risk scores

Kristi Läll MSc, Reedik Mägi PhD, Andrew Morris PhD, Andres Metspalu MD, PhD & Krista Fischer PhD

Genetics in Medicine (2017) 19, 322–329 | doi:10.1038/gim.2016.103

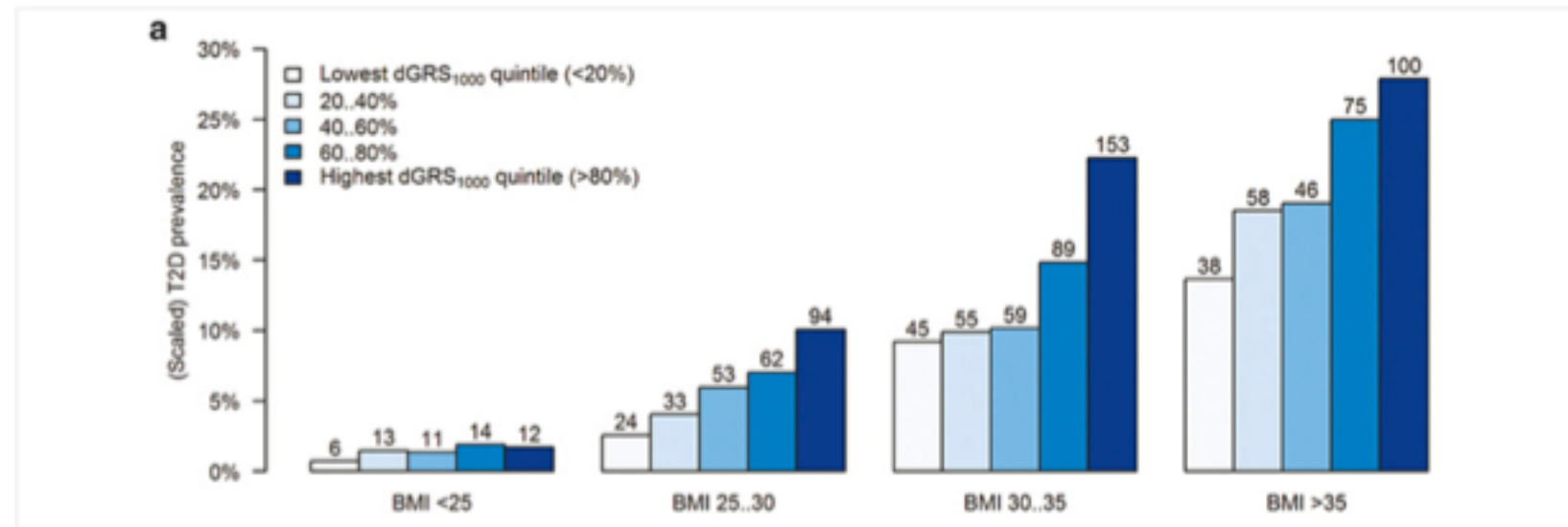


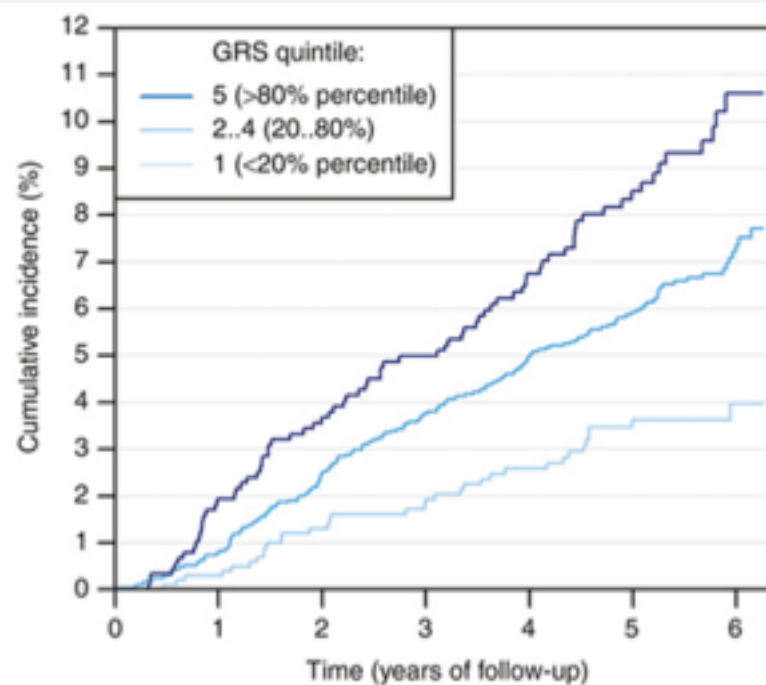
Figure 2

From

Personalized risk prediction for type 2 diabetes: the potential of genetic risk scores

Kristi Läll MSc, Reedik Mägi PhD, Andrew Morris PhD, Andres Metspalu MD, PhD & Krista Fischer PhD

Genetics in Medicine (2017) 19, 322–329 | doi:10.1038/gim.2016.103



Cumulative incidence of type 2 diabetes in 4,881 genotyped individuals free of T2D aged 35–79 and with BMI >23 at baseline. In the figure, 6.25-year follow-up is presented because only 25% of individuals were followed for more than 6.25 years. Cumulative incidence is presented separately in three dGRS₁₀₀₀ categories.

Genetic risk estimation for all gene donors

Sirje Saar

Sinu andmed  Tü - Sirje Saar (MRS)

Mees Naine

Vanus

Kaal Pikkus Vööübermõõt

Hüpertoonia Müskardinfarkt

Geneetilise riski skoor (GRS)

[Arvuta risk](#)

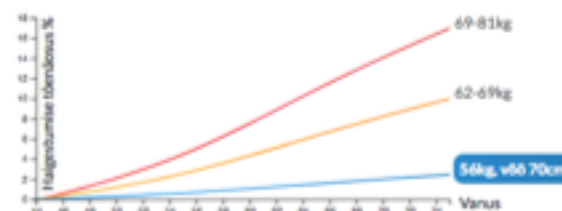
Pärilik diabeedirisk



Sinu geenide poolt määratud diabeedirisk on keskmine. Sinu elustiili poolt määratud risk on madal.

Kokkuvõttes on sinu risk haigestuda elu jooksul teist tüüpi diabeeti madal.

Diabeedirisk sõltub kehakaalust



Sinu tõenäosus haigestuda järgmise 10 aasta jooksul diabeeti on 0.5%. Tõenäosus haigestuda enne 70. eluaastat on 2%.

Pärilik südame-veresoonkonna haiguste risk



Sinu geenide poolt määratud südame-veresoonkonna haiguste risk on kõrge.

Pärilik varajase menopausi risk

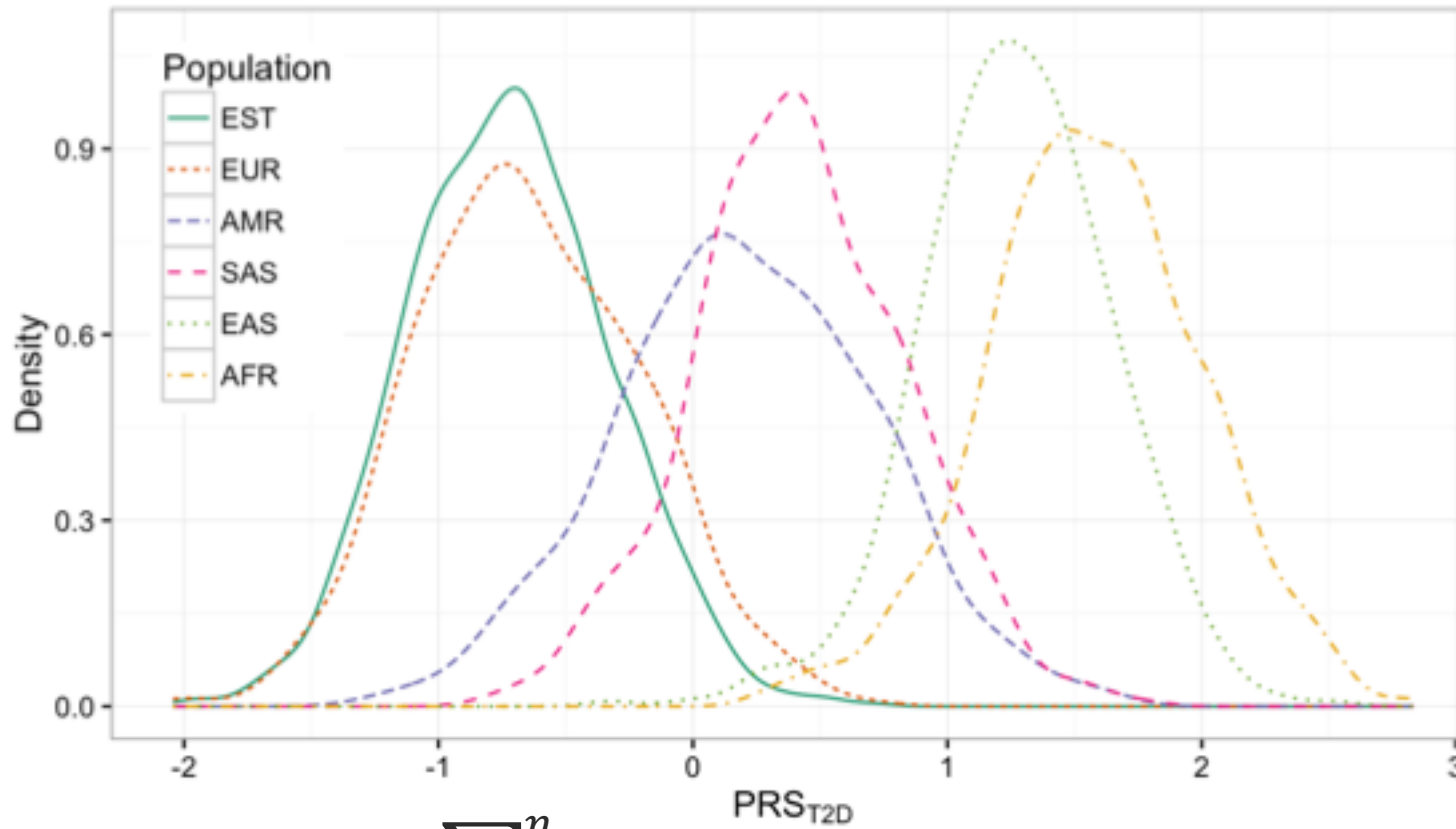


Sinu geenide poolt määratud varajase menopausi risk on kõrge.

Comparing distributions of polygenic risk scores of type 2 diabetes and coronary heart disease within different populations

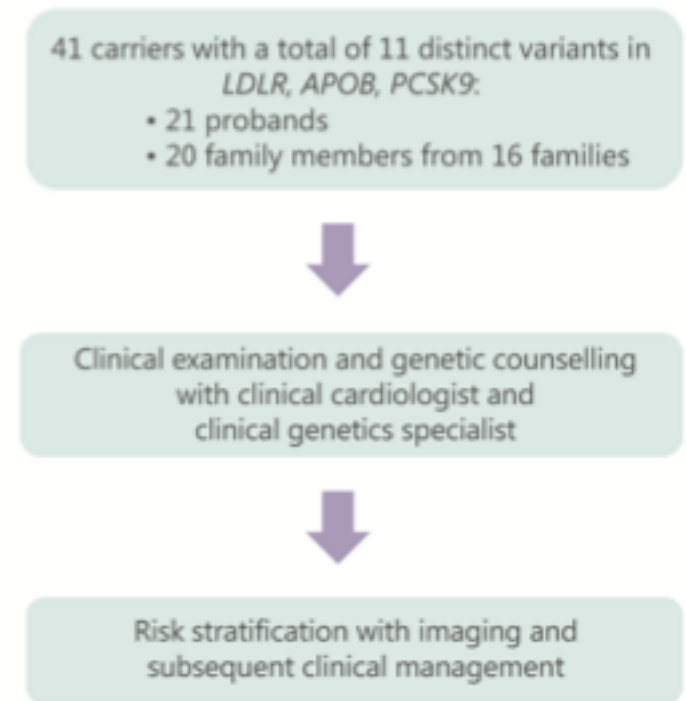
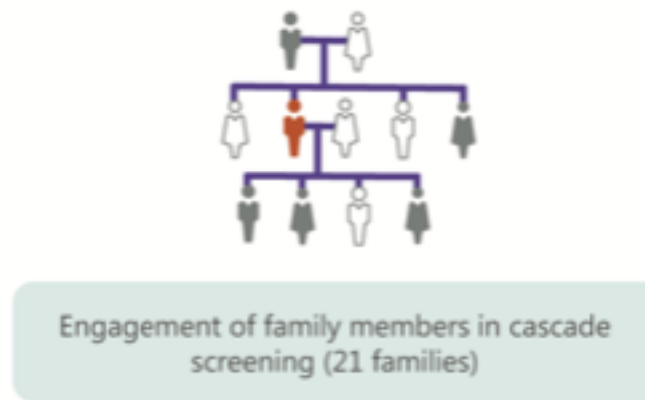
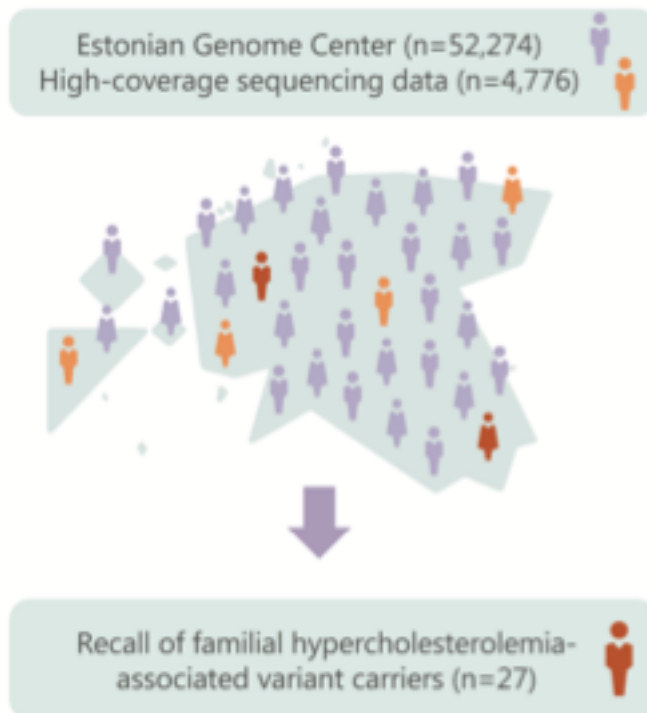
Sulev Reisberg, Tatjana Ijasenko, Kristi Lill, Krista Fischer, Jaak Vilo

Published: July 5, 2017 • <https://doi.org/10.1371/journal.pone.0179238>

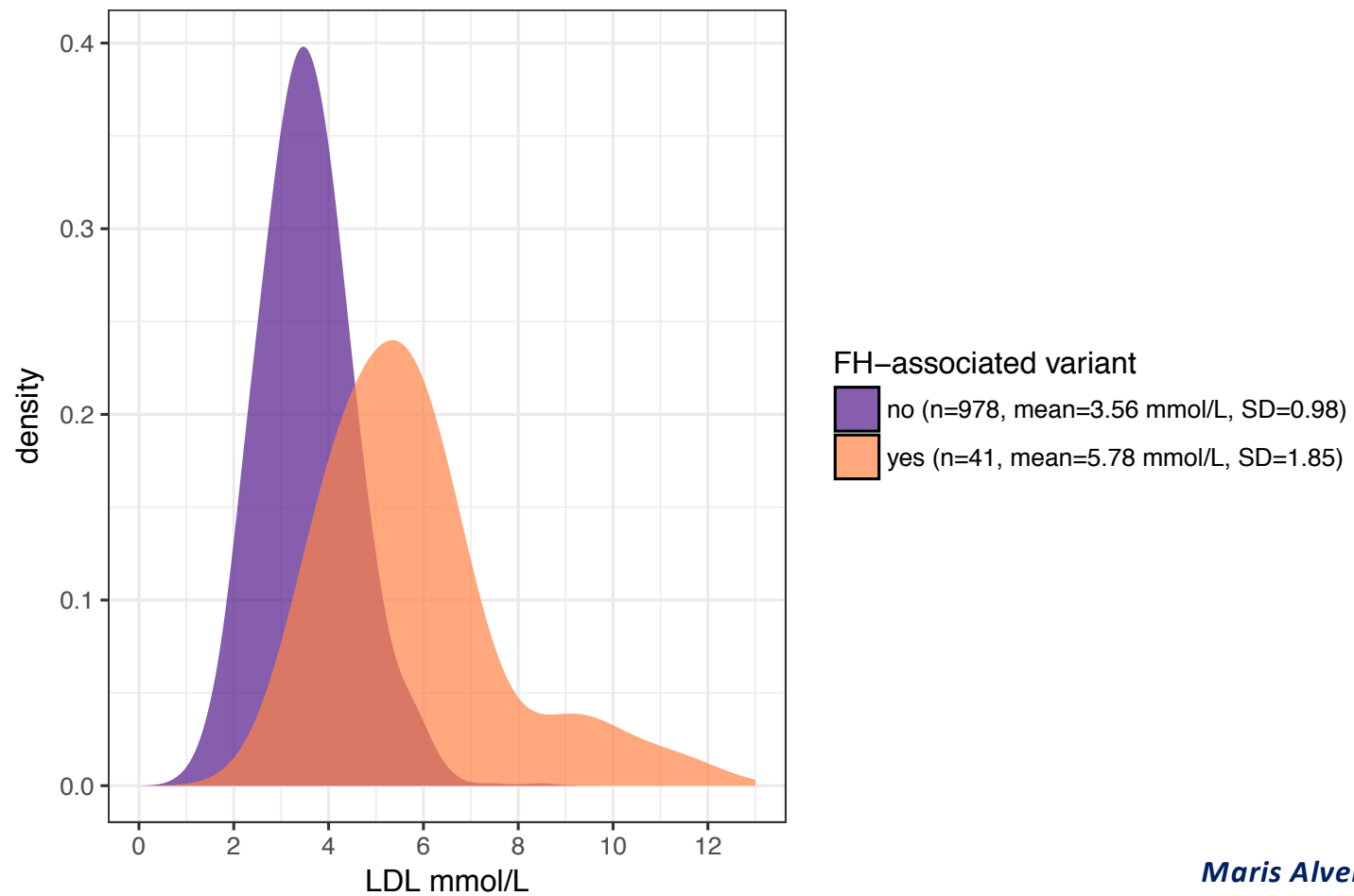


$$PRS = \sum_{i=1}^n w_i \cdot X_i$$

Familial hypercholesterolemia (FH): Recall by genotype and cascade screening

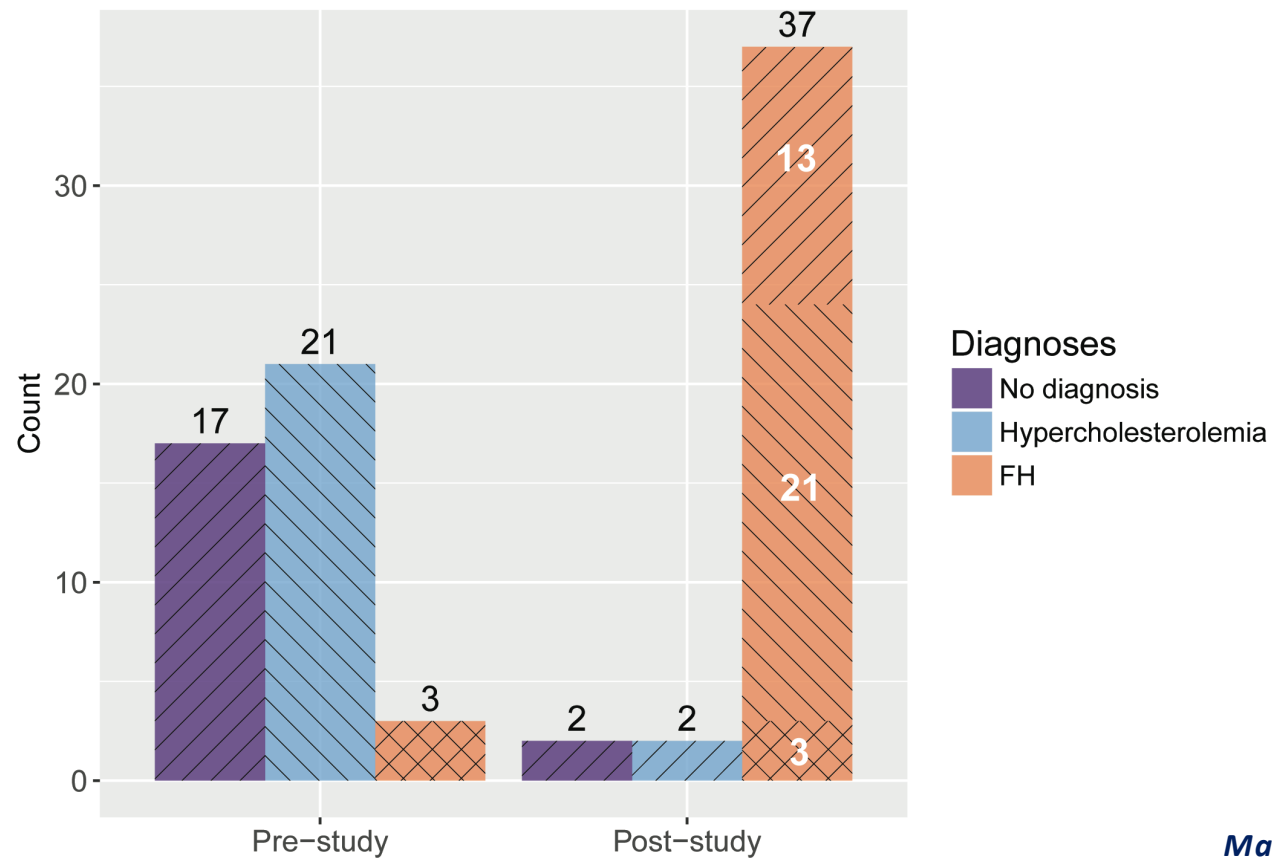


FH variant carriers display a wide spectrum of LDL concentrations, overall increased by 2.29 mmol/L



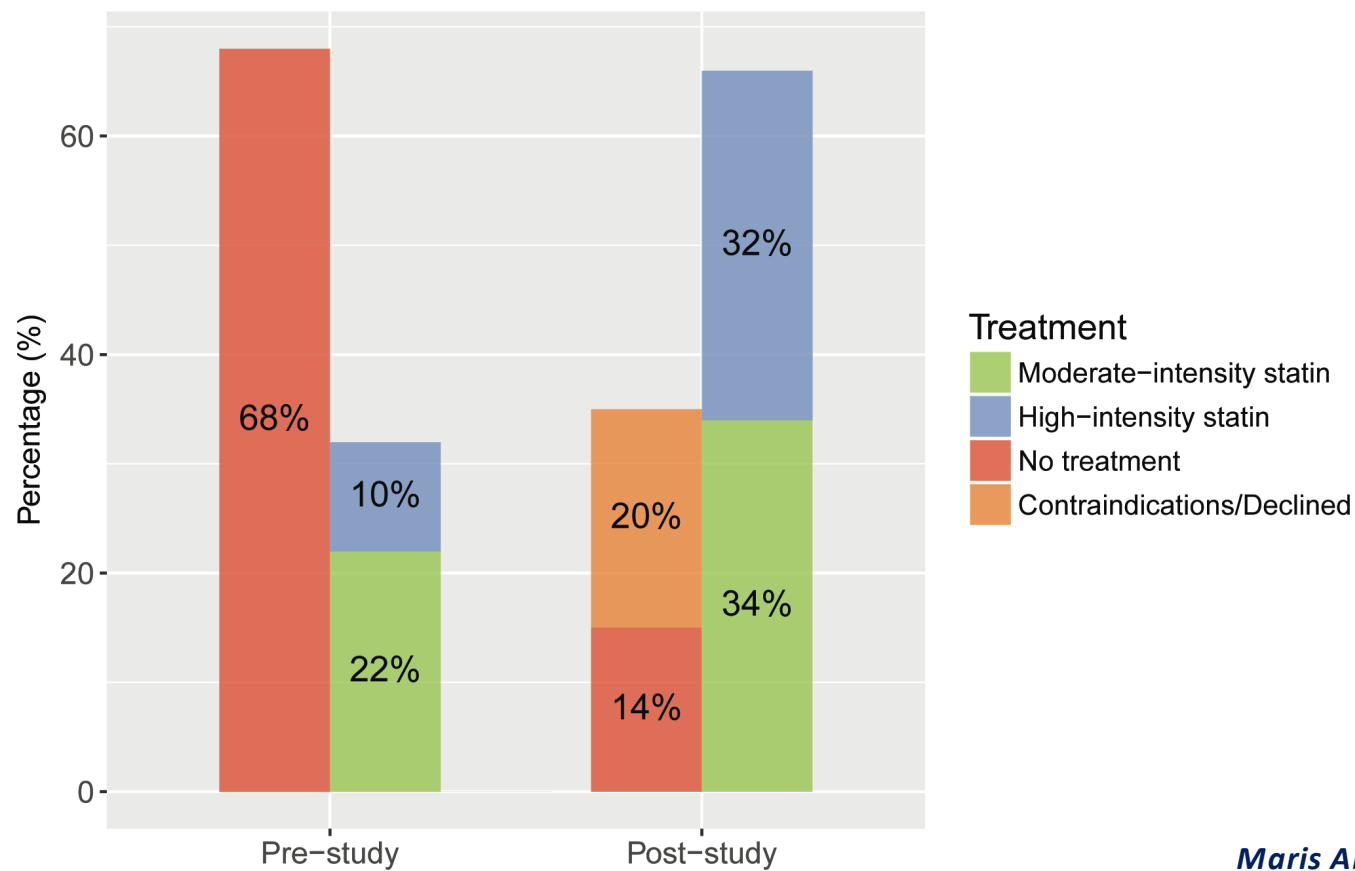
Maris Alver, Dr Tõnu Esko

Disease diagnoses in FH-associated variant carriers

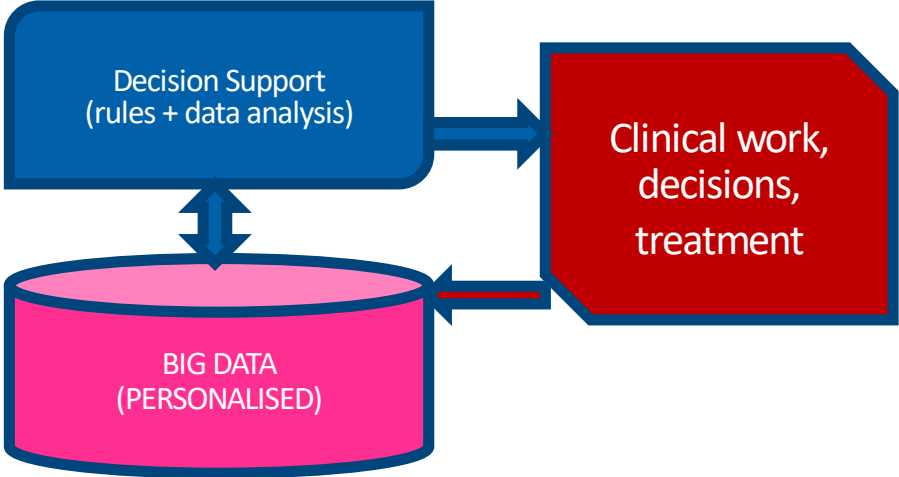


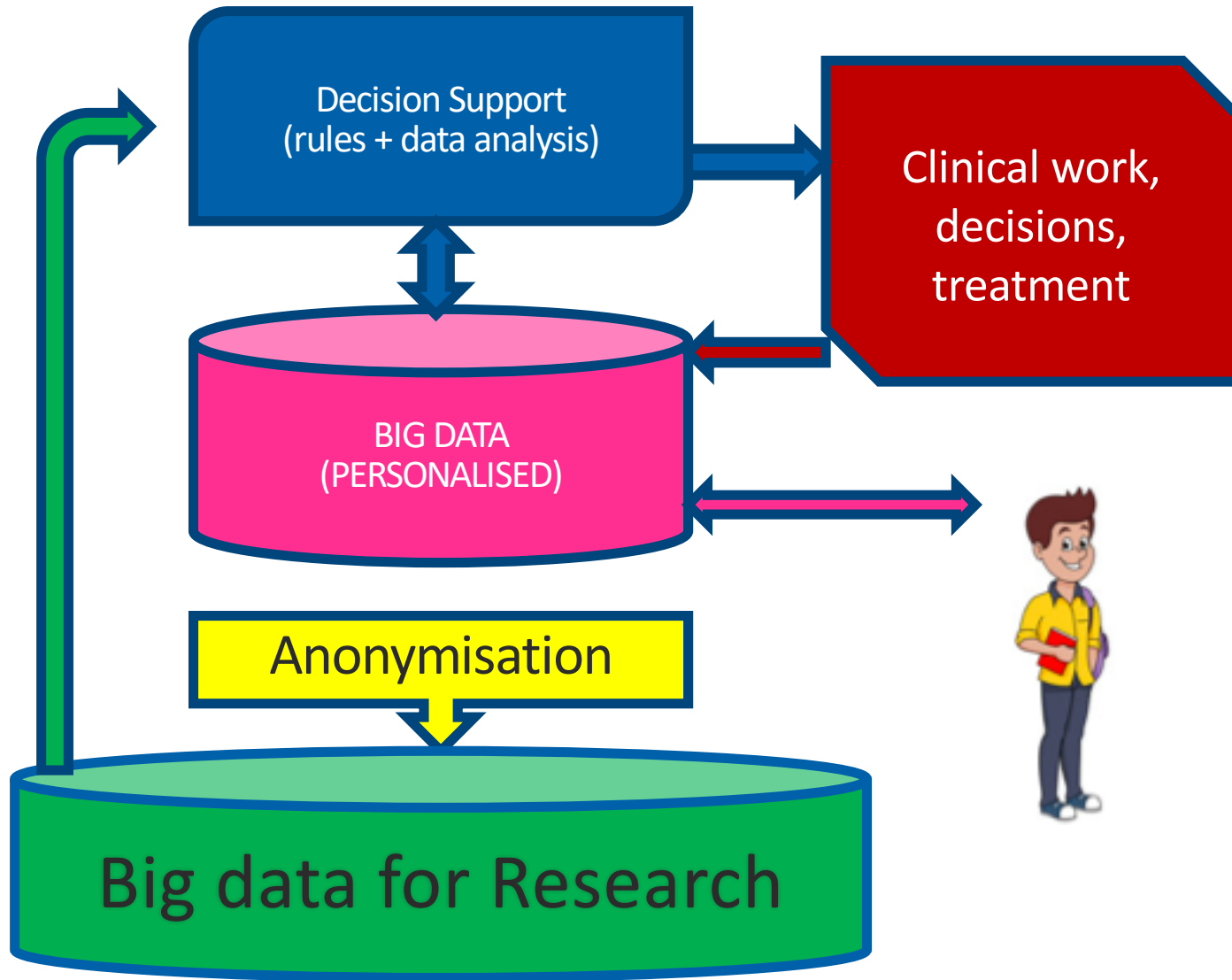
Maris Alver, Dr Tõnu Esko

Statin treatment in FH-associated variant carriers

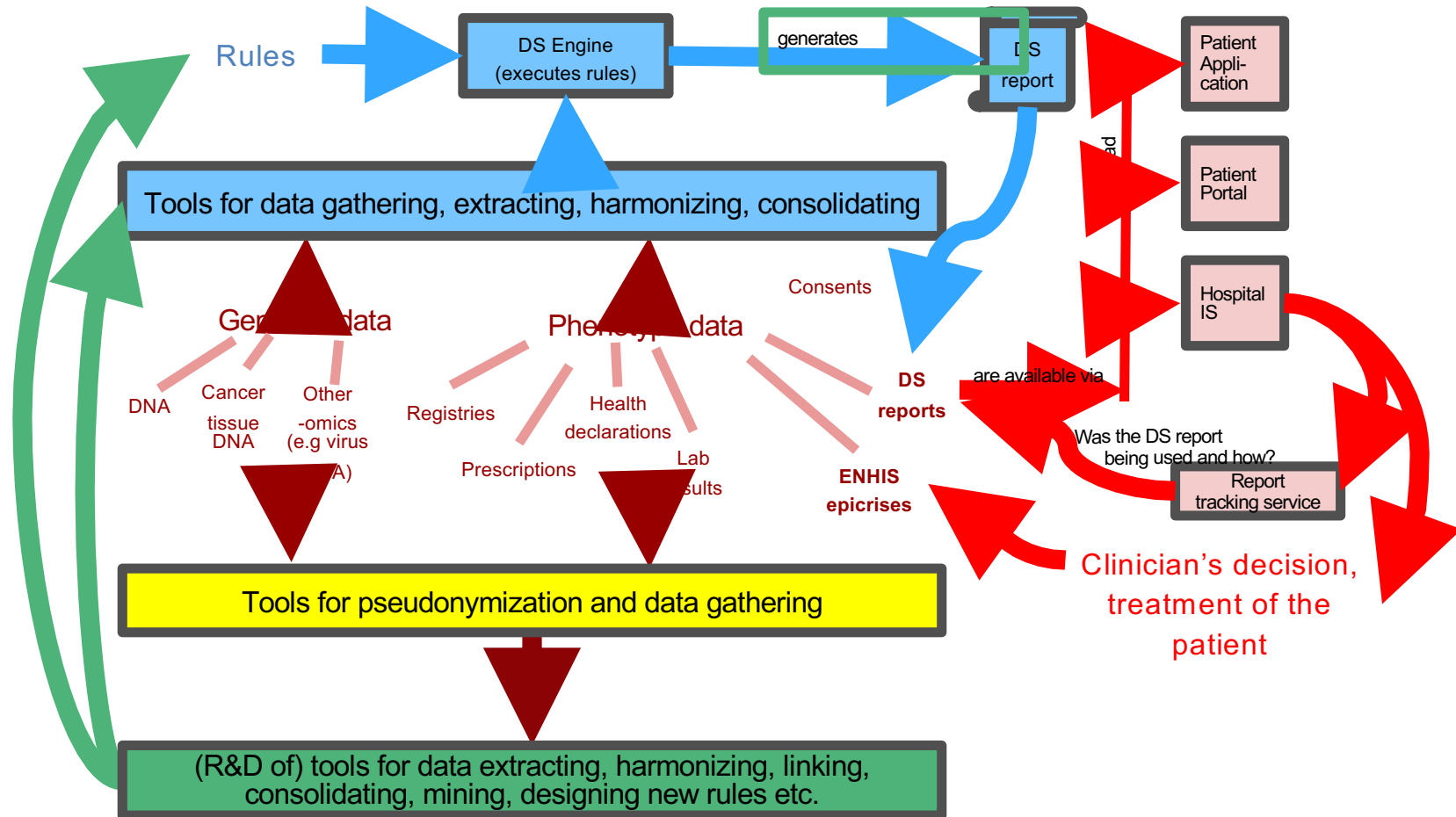


Maris Alver, Dr Tõnu Esko





What tools & data in each process



Research and innovation challenges

- **Collecting**, cleaning, systematising national health data
- **Linking** different databases by pseudonyms
- **Omics data readiness** across the field
- **Secure data handling** environments
- **Decision Support Systems** for healthcare
- **Self-reporting and wearables**
- **International** comparison, analysis, benchmarking



Prof. Andres Metspalu, Lili Milani, Tõnu Esko, Krista Fischer, Reedik Mägi, Maris Alver, Kristi Läll, Kristi Krebs, Tõnis Tasa, Mart Kals, Tom Haller, Neeme Tõnisson, Anu Reigo, Liis Leitsalu, Helene Alavere, Kristjan Metsalu, Kairit Mikkel, Mari-Liis Tammesoo



Prof. Jaak Vilo, Hedi Peterson, Sulev Reisberg, Sven Laur, jt.



BioMed@STACC: the health data mining team

▼ 18 people



- Alexander Tkachenko
- Anne-Liis Tanav
- Dage Särg
- Gea Pajula
- Harry-Anton
- Karl-Oskar Masing
- Marek Oja
- Margus Jäger
- marje johanson
- Markus Lippus
- Mihkel Manguse
- monika soosaar
- Raul Sirel
- Robert Roosalu
- Sulev Reisberg
- Sven Laur
- Tormi Reinson
- Uku R



REPUBLIC OF ESTONIA
MINISTRY OF SOCIAL AFFAIRS



Tervise Arengu Instituut
National Institute for Health Development

Dr Silja-Riin Voolma, Dr Priit Tohver

Dr Annika Veimer, Dr Toomas Veidebaum



Prof. Eric Lander, Dr Sekar Kathiresan,
Dr Daniel McArthur





STACC Software Technology and Applications Competence Center



BBMRI-ERIC

Biobanking and
BioMolecular resources
Research Infrastructure



Thank you!
Jaak Vilo
vilo@ut.ee

Contact:

Jaak Vilo

UniTartuCS

Head of Institute

vilo@ut.ee



UNIVERSITY OF TARTU
Institute of Computer Science

