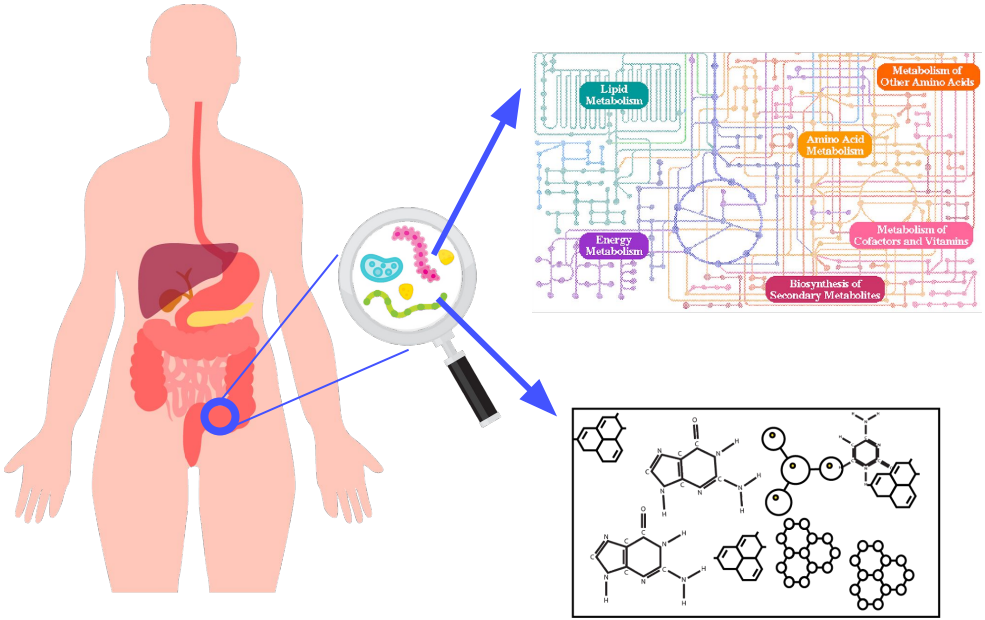
Heart Failure Prediction: Microbiome
FINRISK DREAM Challenge

Pande Putu Erawijantari
Turku Collegium for Science, Medicine, and Technology – Postdoctoral researcher
Department of Computing, Faculty of Technology, University of Turku,Finland

pande.erawijantari@utu.fi @erawijantaript

ISMB ECCB 2023 LYON FRANCE July 23-27 ISCB

Overview of human gut microbiome



Factors

- Diet
- Lifestyle
- Geography
- Age
- Medication
- Diseases
- Genetics

Roles

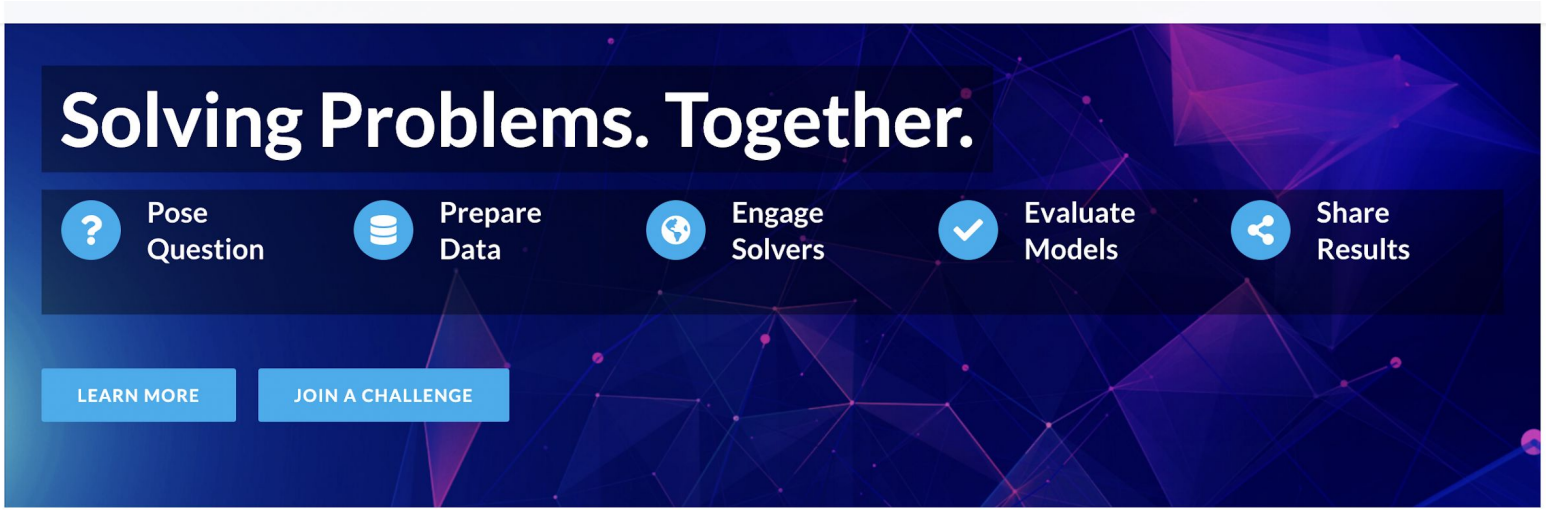
- Protecting host against pathogenic microbes
- Modulating immunity
- Regulating metabolic processes

Li et al., Nature Biotechnology, 2014, Human body picture designed by freepik.com

UNIVERSITY OF TURKU

DREAM CHALLENGES
powered by Sage Bionetworks

What is DREAM Challenge?

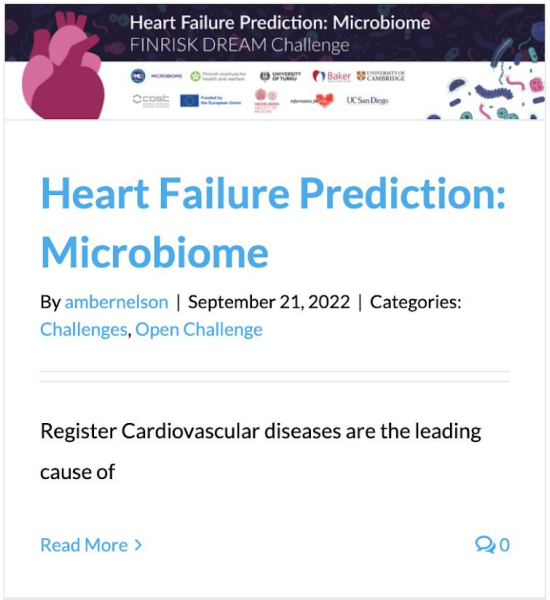


DREAM Challenges use crowd-sourcing to solve complex biomedical research questions

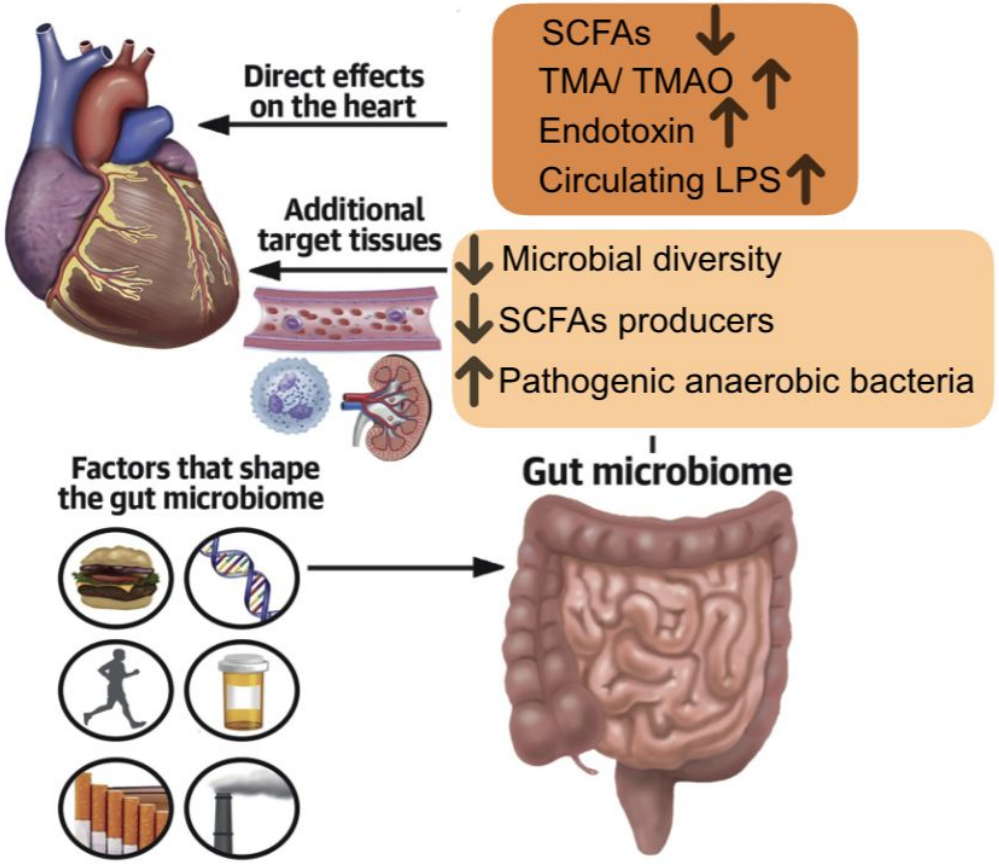
Open Challenges

Scan for more info

https://synapse.org/finrisk



Screenshot was taken on the challenge launch in September, 2022

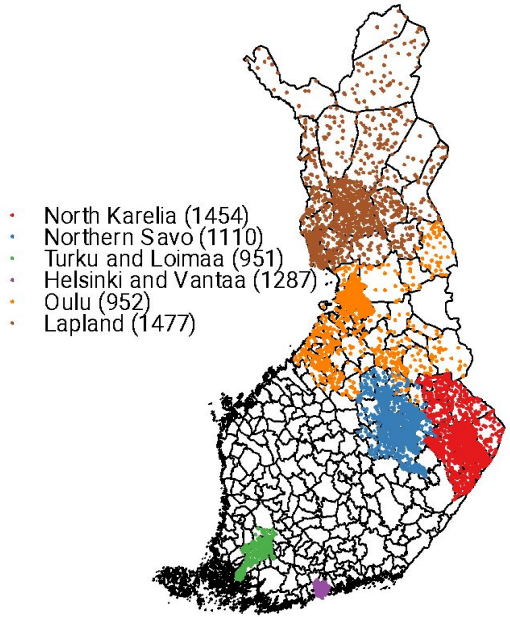


Can microbiome improve the predictions of heart failure?

Adapted from Mamic et al., 2023

The National FINRISK Study - 2002

- Finnish prospective population cohort study
- Individual were followed through for 15+ years.
- There are 559 incident heart failure among 7231 participants



Collected and hosted by

Finnish institute for health and welfare

13,498 invitees

8,738 participated

7,231 donated Fecal sample (54% participation rate)

Sequence by UC San Diego

DNA extractions And shallow metagenomic sequencing (~1M reads/sample)

Borodulin et al., 2018

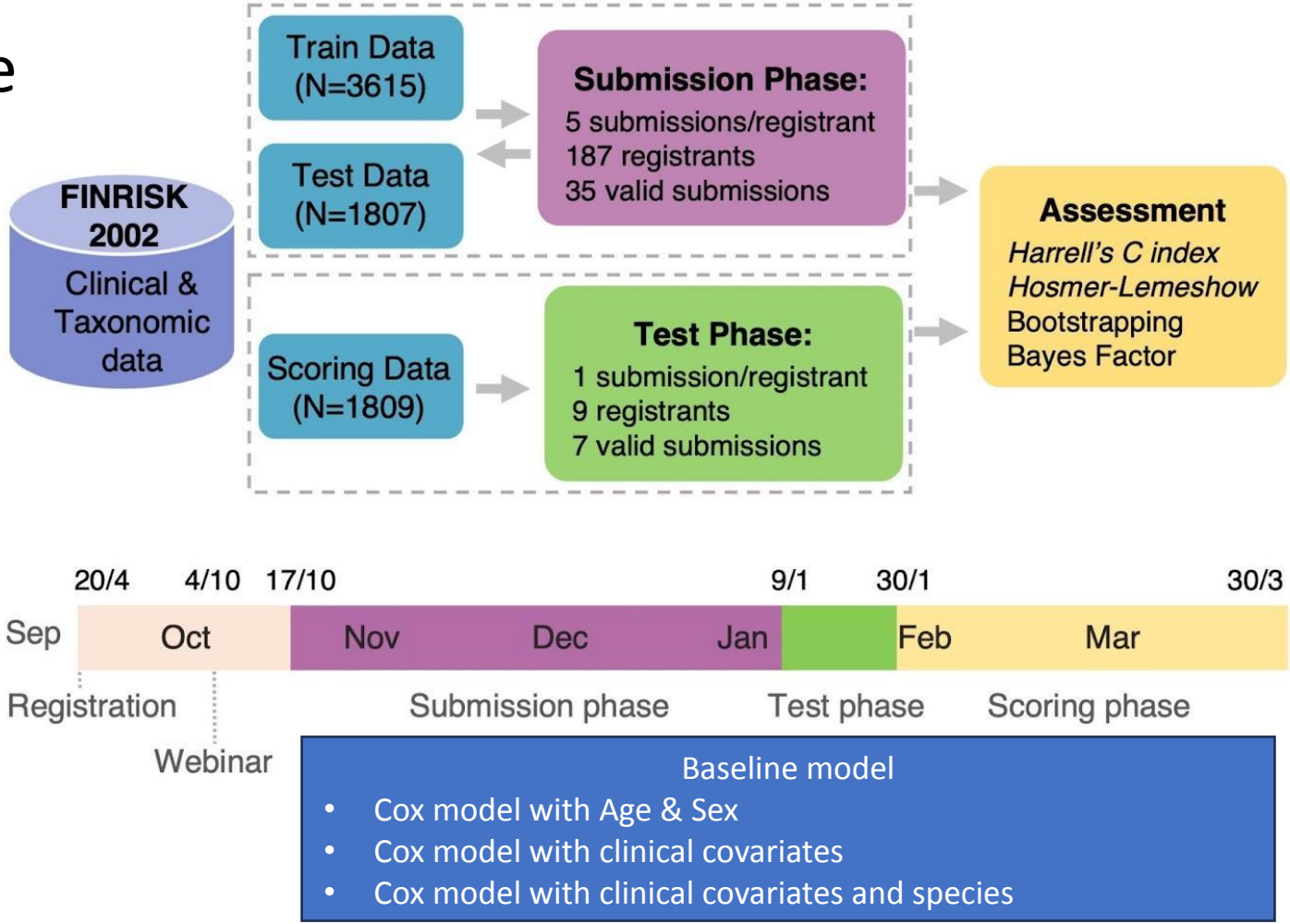
Highlight of the FINRISK 2002 microbiome studies

- Taxonomic signatures of cause-specific mortality risk in human gut microbiome (Salosensaari et al., 2022)
- Identifications and characterizations of combined effects of host genetics and diet on human gut microbiota and incident diseases (Qin et al., 2022)
- Gut microbiome signatures are predictive of incident type 2 diabetes (Ruuskanen M., **Erawijantari PP.**, et al., 2022)

Learn more from https://datascience.utu.fi/

@openreslabs

Challenge overview



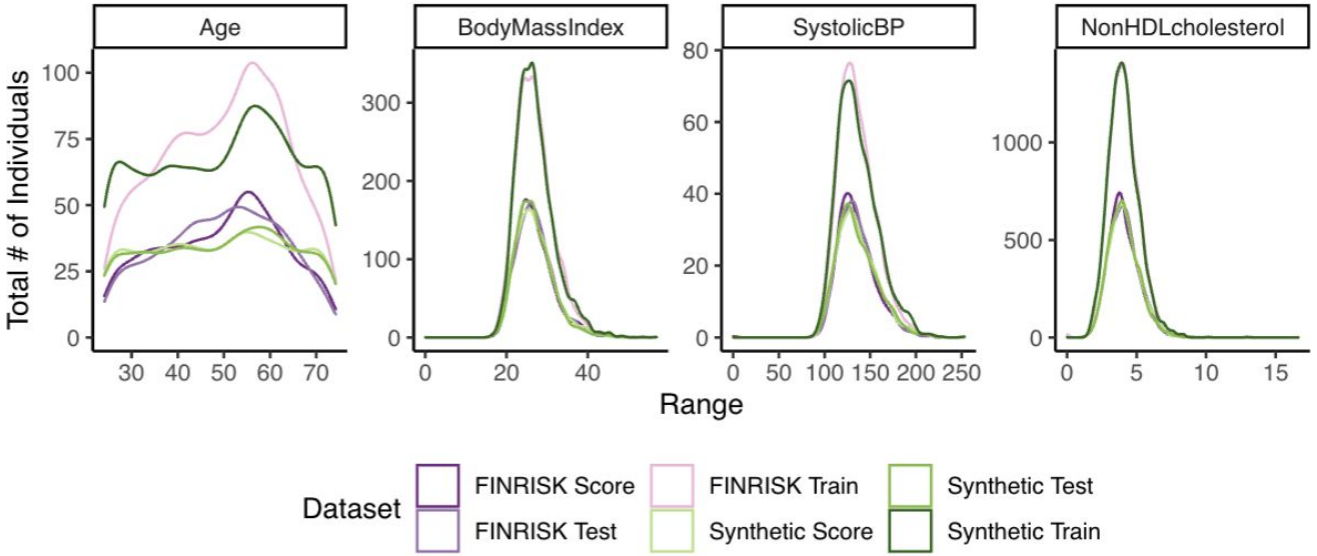
Synthetic data generation: Why is it necessary?

IMPORTANT! The FINRISK data contains sensitive personal information from healthcare registers which cannot be shared **without formal agreement with THL biobank**

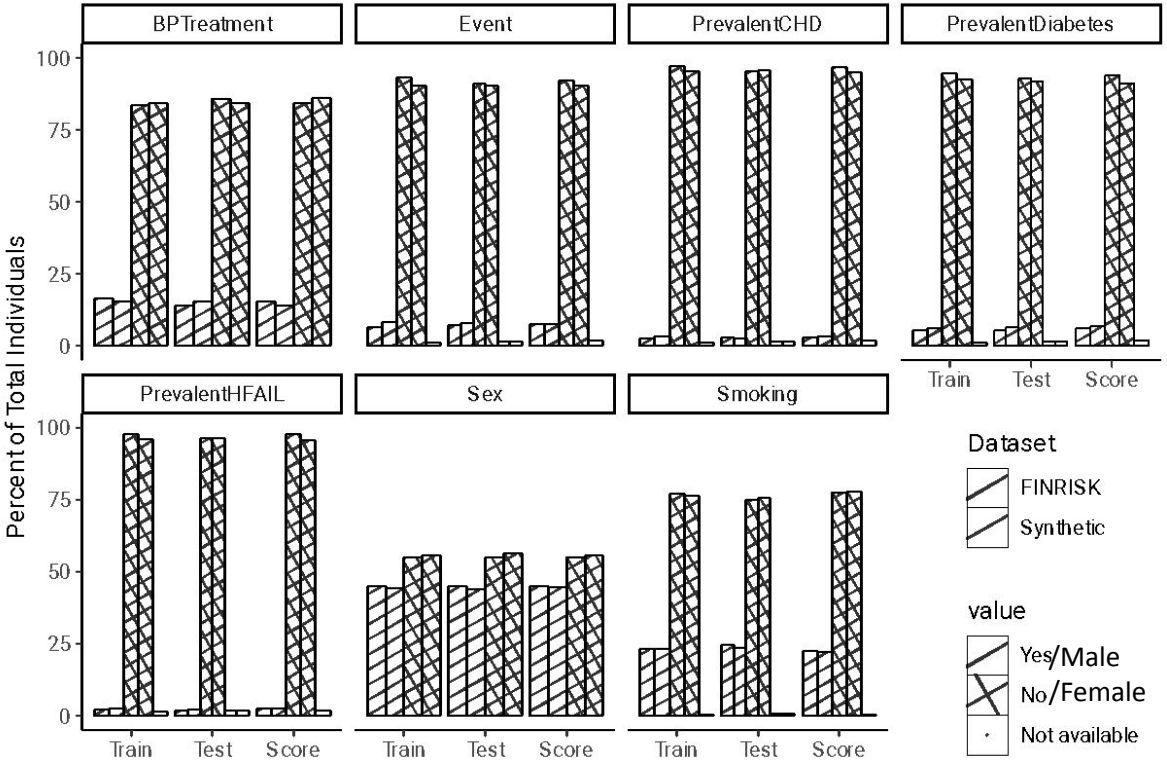
Criteria: preserve the covariance structure while privatizing the data

Rank-based Inverse Normal Transformation (by Aki Havulinna, 2022)

Synthetic data – continuous covariates



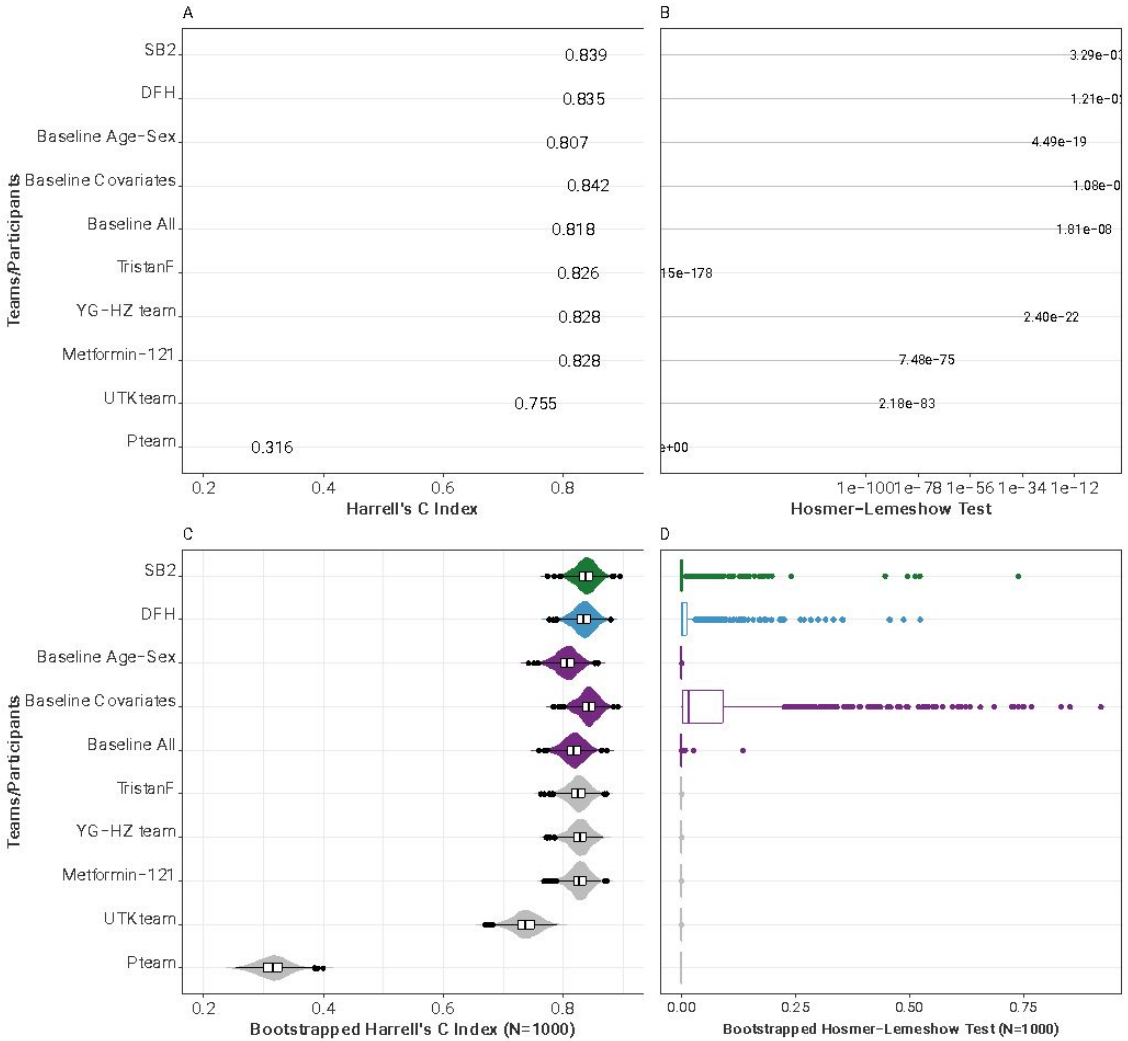
Synthetic data – categorical covariates



Received submissions

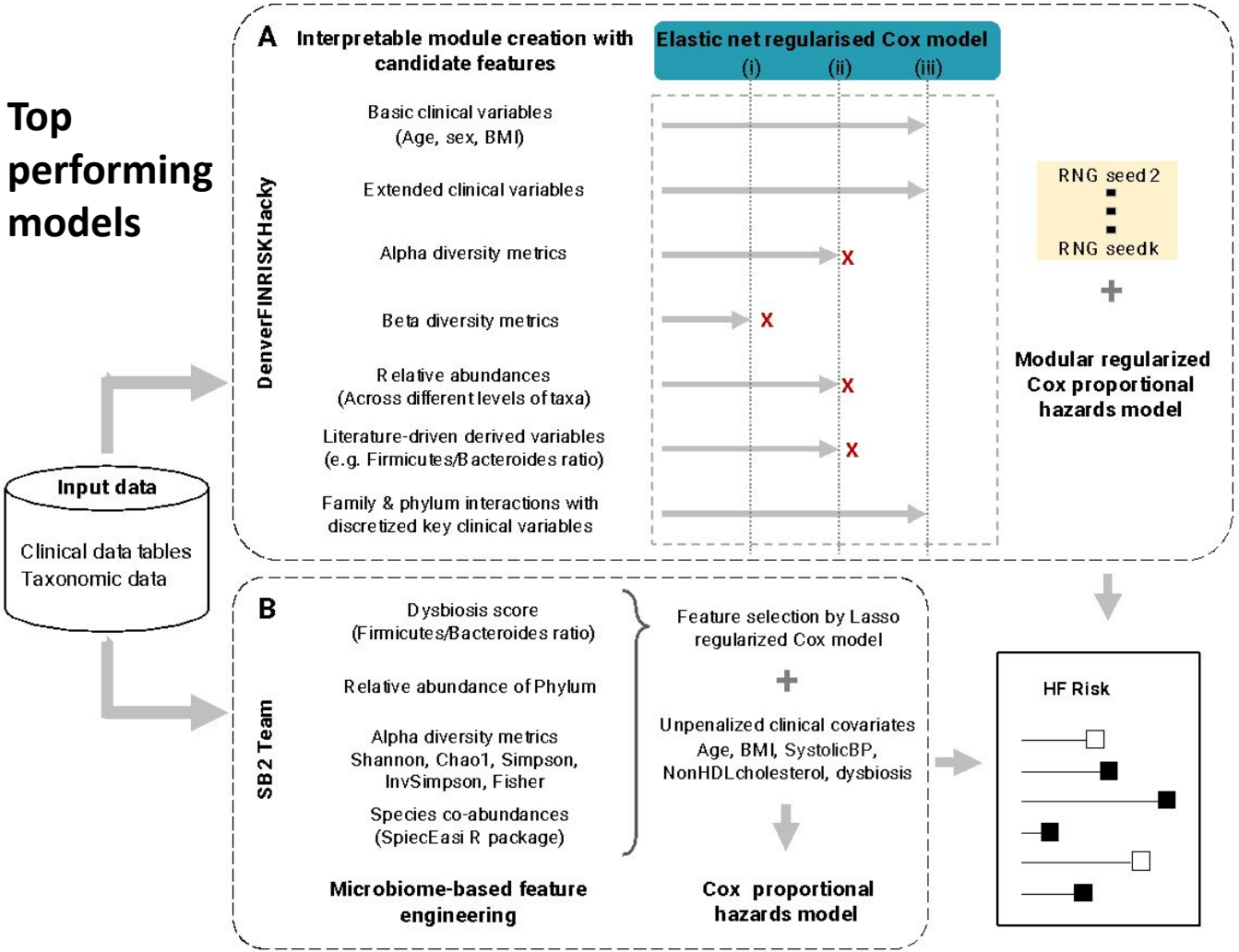
- 35 VALID submissions from 9 participants were accepted in submission phase
- 7 teams submitted final model

id	createdOn	submitterid	status	harrell_c	hoslem_test
9731637	1/29/2023 11:29 PM	SB2	SCORED	**0.839395379197954**	0.0032924949185753
9731625	1/29/2023 7:12 PM	DenverFINRISKHacky	SCORED	**0.835136923343354**	0.0120891360271361
9731490	1/27/2023 1:26 AM		SCORED	**0.827974369335565**	2.40190686358651e-22
9731454	1/26/2023 11:55 AM		SCORED	**0.827752343482875**	7.48389464769304e-75
9731636	1/29/2023 11:02 PM		SCORED	**0.825842921149739**	1.14999761860238e-178
9731713	1/31/2023 12:33 AM		SCORED	**0.754830172425277**	2.18334641086665e-83
9731666	1/30/2023 9:35 PM		SCORED	**0.315609749599243**	0



Received model performance vs Baseline model

Top performing models

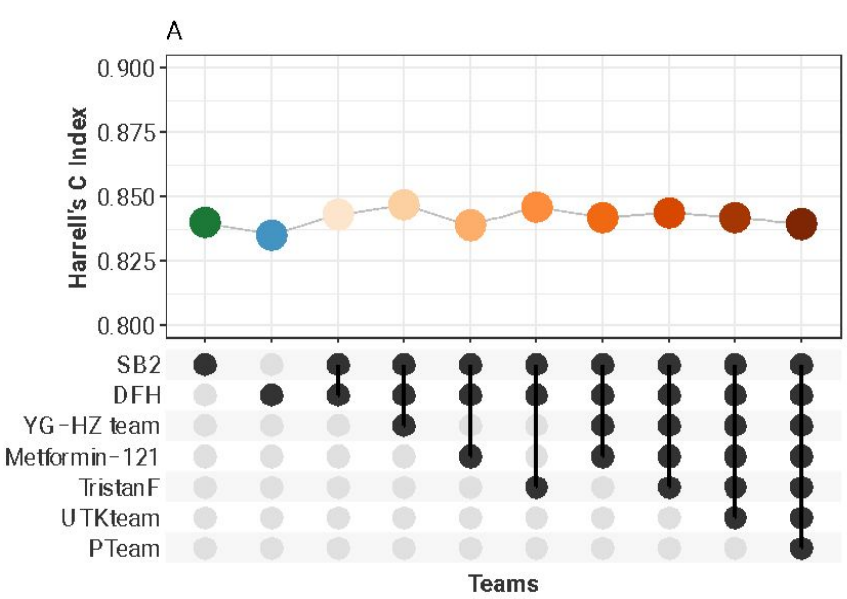


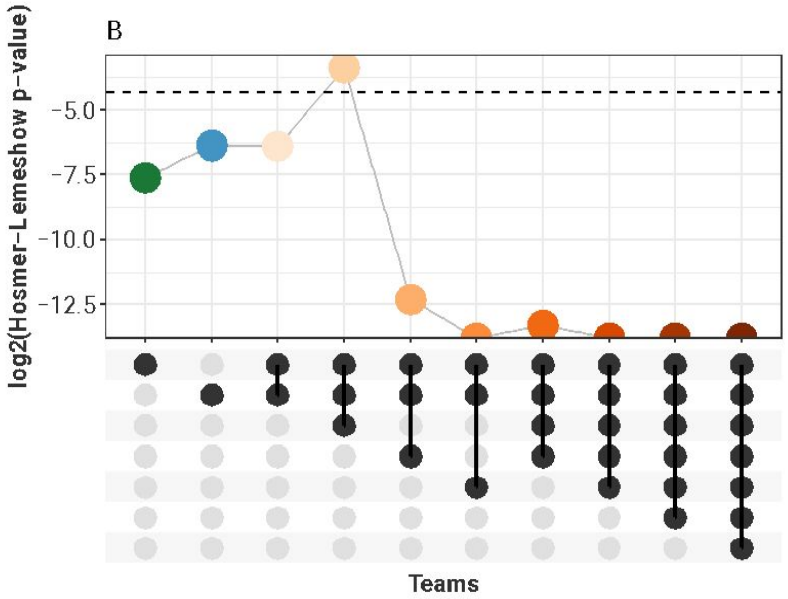
https://www.synapse.org/#!Synapse:syn50613972/wiki/620396

https://www.synapse.org/#!Synapse:syn47950082/wiki/620639

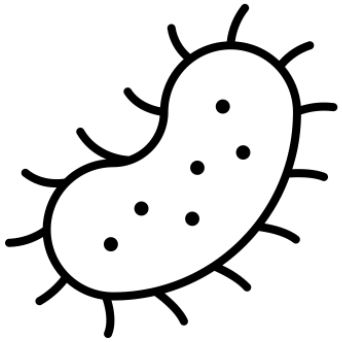
Post Challenge Phase

- Model refinement
- Model ensemble
- Investigating the most predictive features to discuss the clinical relevancy of the model





Microbiome perspective in HF-predictions

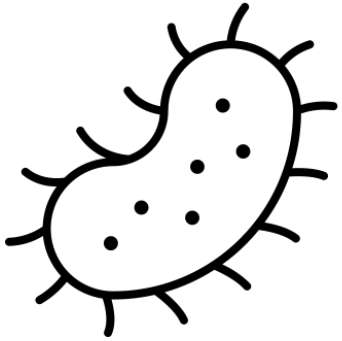


- Clinical covariates are stronger predictors for HF and microbiome features offer supplementary predictive value to improve model performance

- Microbiome-related features represent:
  - negative association of alpha-diversity index with incident HF
  - **Inflammation signatures**: importance of species cluster consisting *R. gnavus*, *C. bolteae*, *C. citroniae*, *C. difficile*)
  - **TMAO signatures**: positive associations with species cluster consisting of *C. citroniae*, *C. asparagiforme*, *H. hathewayi*)

Perspective of the challenge

- Open avenue to explore the potential of microbiome-based biomarkers to complement clinical risk factors in predicting individuals with an elevated risk of HF
- The crowdsourced *Challenge* provides a unique platform for collaboratively solving scientific problem
- Highlight the need for further investigation into model ensemble approaches to enhance the overall predictive capability and clinical relevance
- The synthetic dataset has proven beneficial for challenge participants and could become an option for data sharing while ensuring privacy protection.



Acknowledgements

Initiated by:

MICROBIOME

COST Action (CA18131)
Action chair: Marcus Claesson, University College Cork, Ireland
WG2 leader: Leo Lahti, University of Turku, Finland

Organized by:

UNIVERSITY OF TURKU

Pande Putu Erawijantari
Rajesh Shigdel
Teemu Niiranen
Leo Lahti

Finnish institute for health and welfare

Aki Havulinna
Pekka Jousilahti
Veikko Salomaa

HEIDELBERG FACULTY OF MEDICINE

Ece Kartal
Rebecca Levinson
Julio Saez-Rodriguez

UC San Diego

Rob Knight
Mohit Jain

SageBionetworks

Funded by:

Informatics for Life

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the European Union

Thanks to

- All participants of the FINRISK 2002
- FINRISK DREAM Challenge participants especially **SB2** team lead by **José Liñares Blanco** and **DenverFINRISKHacky** team lead by **Teemu D. Laajala** who has been active in the post-challenge phase
- Turku Data Science Group, Department of Computing, University of Turku, Finland, twitter: **@openreslabs**

CSC

- Computing support from CSC (Finnish IT Center for Scientific Computing)

Thank you

Q & A

Visit our poster number B-190
by Ece Kartal

Predicting Incident Heart Failure from the Gut Microbiome: The DREAM FINRISK challenge

Authors: Ece Kartal[1*], Pande Putu Erawijantari[2*], José Liñares Blanco[3,4], Teemu D. Laajala[5,6], Lily Elizabeth Feldman[7], The FINRISK Microbiome DREAM Challenge Community, Pedro Carmona Saez[8,9], Rajesh Shigdel[1], Michael Inouye[5,6], Pekka Jousilahti[10], Mohit Jain[11,12], Rob Knight[13,14,15,16], Veikko Salomaa[10], Teemu Niiranen[10,17,18], Aki S. Havulinna[10,19], Julio Saez-Rodriguez[4], Rebecca T. Levinson[7,20], Leo Lahti[2+]
+shared first; *shar ed last Contact: leo.lahti@iki.fi; rebeccaterrell.levinson@med.uni-heidelberg.de; ece.kartal@uni-heidelberg.de @ps_ecekartal; saezlab

1. Motivation and Aim

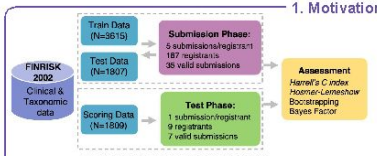


Fig 1. Dream FINRISK Challenge Overview

Heart failure (HF) is a complex clinical syndrome characterized by the heart's inability to meet the body's blood supply needs. Several studies have foundr diffe ences in the microbiome composition of HF patients compared to controls.

The crowdsourced **FINRISK Microbiome DREAM challenge** (**Fig 1**) aimed to investigate the **gut microbiome compositions in predicting HF risk** in a large population of 7,231 Finnish adults' (FINRISK 2002, n = 559/7,231 HF). To protect the privacy of individuals, we provided synthetic data that closely mimics the real FINRISK data.

2. Modelling Workflow

Two teams **(SB2, DFH)** developed superior models that outperformed the baseline and other competing models using regression-based methods with diffe ent approaches for feature importance (**Fig 2**).

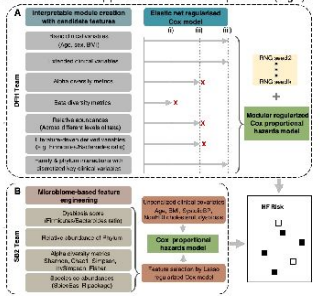


Fig 2. Modeling Workflo of the Top Two Performing Teams

3. Model Results

- Baseline Age-Sex: Cox model with only Age + Sex as covariates
- Baseline Covariates: Cox model with all clinical covariates
- Baseline All: Cox model with all clinical covariates + microbiome

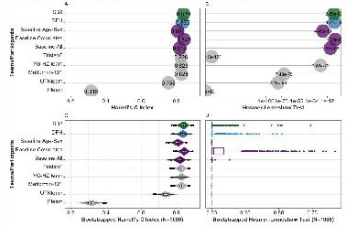


Fig 3. Harrell's C-index and Hosmer-Lemeshow test results

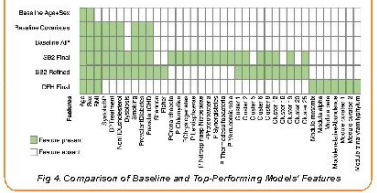


Fig 4. Comparison of Baseline and Top-Performing Models' Features

4. Model Refinement

Post-challenge enhancements included minimizing redundant taxonomic information through species and network analysis, along with combining the finl re thods (**Fig. 5**)

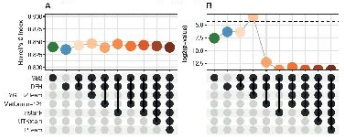


Fig 5. The ensemble model's Harrell's C-index and Hosmer-Lemeshow p-values

5. Conclusion

Microbiome species, particularly those **associated with inflamma tion** , were found to be valuable predictors of incident HF.

This challenge provides a ground for the scientific community to contribute and advance our current understanding of the **incidence of heart failure and its associations with the gut microbiome.**

Learn More

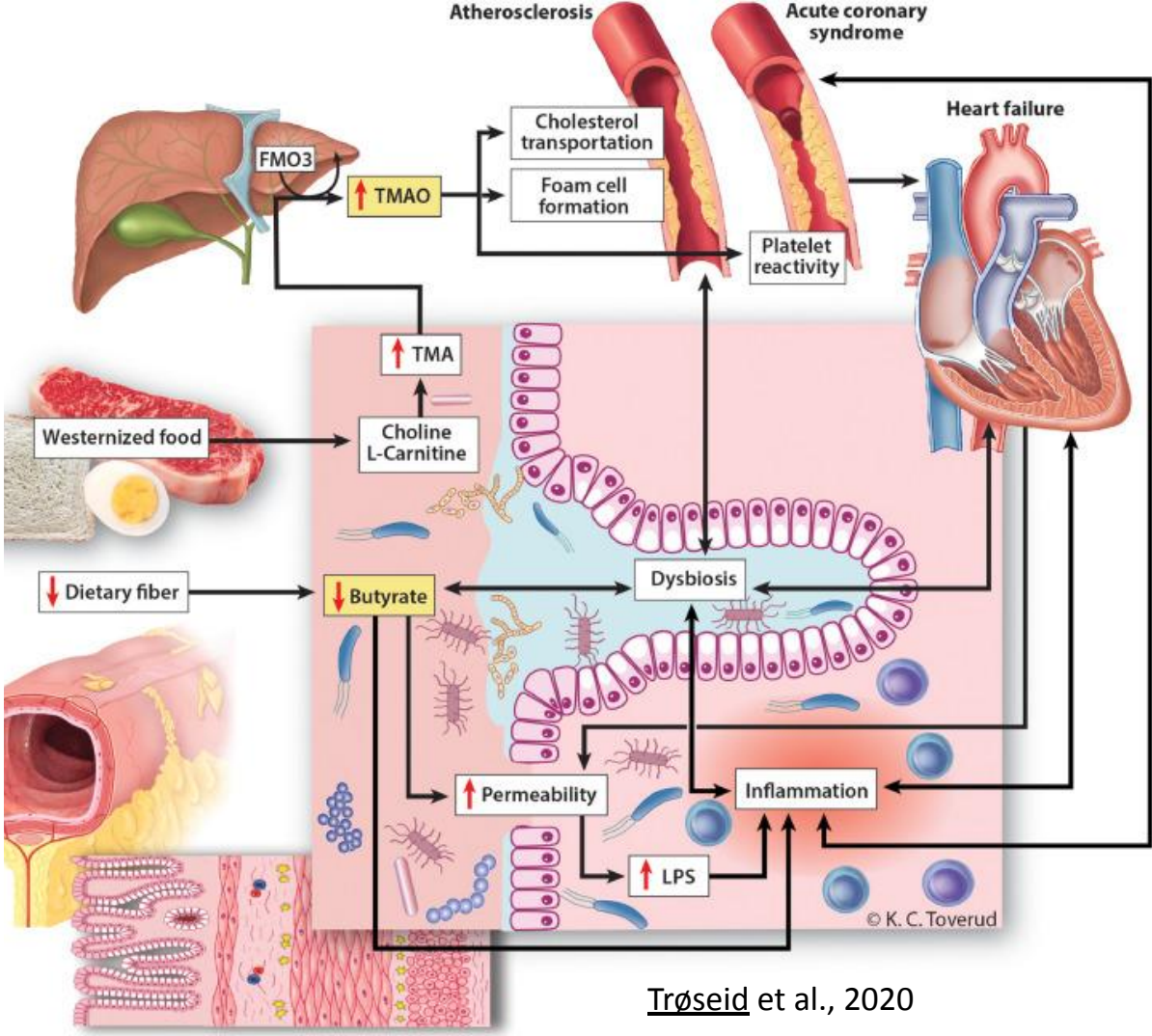
The DREAM challenge main organizer:
COST action network ML4microbiome (CA18131; ml4microbiome.eu)
Challenge page > synapse.org/finisk

References

Salosensaari A, Laitinen V, Havulinna AS et al., Taxonomic signatures of cause-specific mortality risk in human gut microbiome. Nat Commun 12, 2671 (2021). https://doi.org/10.1038/s41467-021-22962-y

Can microbiome improve the predictions of heart failure?

- Heart failure remains difficult to diagnose due to the heterogeneity of the disease and a lack of agreement of diagnostic criteria
- The link between heart failure and the microbiome has long been postulated.
- Lack of study with temporal follow-up

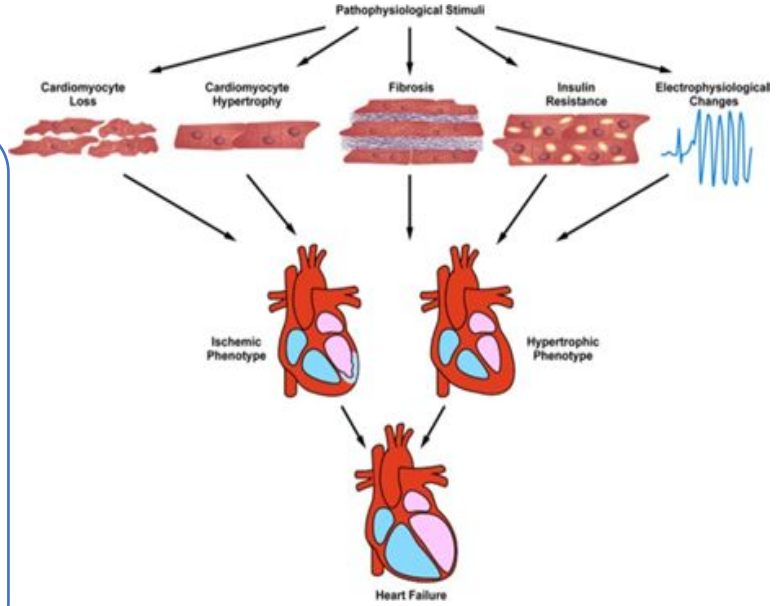


Trøseid et al., 2020

What is Heart failure (HF)

- Heart failure is a clinical syndrome with signs and symptoms caused by structural and/or functional cardiac abnormality with elevated natriuretic peptide levels and/or evidence of pulmonary or systemic congestion.

- Heart failure occurs when the heart cannot pump blood to meet the needs of the body normally.



Definition adapted from Bozkurt et al.. J Card Fail. 2021 Mar 1:S1071-9164(21)00050-6. PMID: 33663906.

Burchfield et al. Circulation. 2013;128:388–400

Bootstrap analysis (n=1000) for robust ranking

1. The individual scores and real survival value in the same order were bootstrapped for 1000 times
2. For each boots, Harrel's C and Hosmer-lemeshow test were performed
3. Bayes Factor was calculated based on the

$$BF(T_1, T_2) = \frac{\sum_{i=1}^{1000} \mathbf{1}(r(T_1)_i > r(T_2)_i)}{\sum_{i=1}^{1000} \mathbf{1}(r(T_1)_i < r(T_2)_i)},$$

Compares the number of times Team 1 is ranked superior to Team 2 with the alternative scenario where Team 2 performs better than Team 1.

National FINRISK study

25 to 74 years old Finnish permanent resident across different geographical area of Finland

years								
1972	1977	1982	1987	1992	1997	**2002**	2007	2012
13500	11507	11395	7932	7927	11500	**13498**	12000	10000

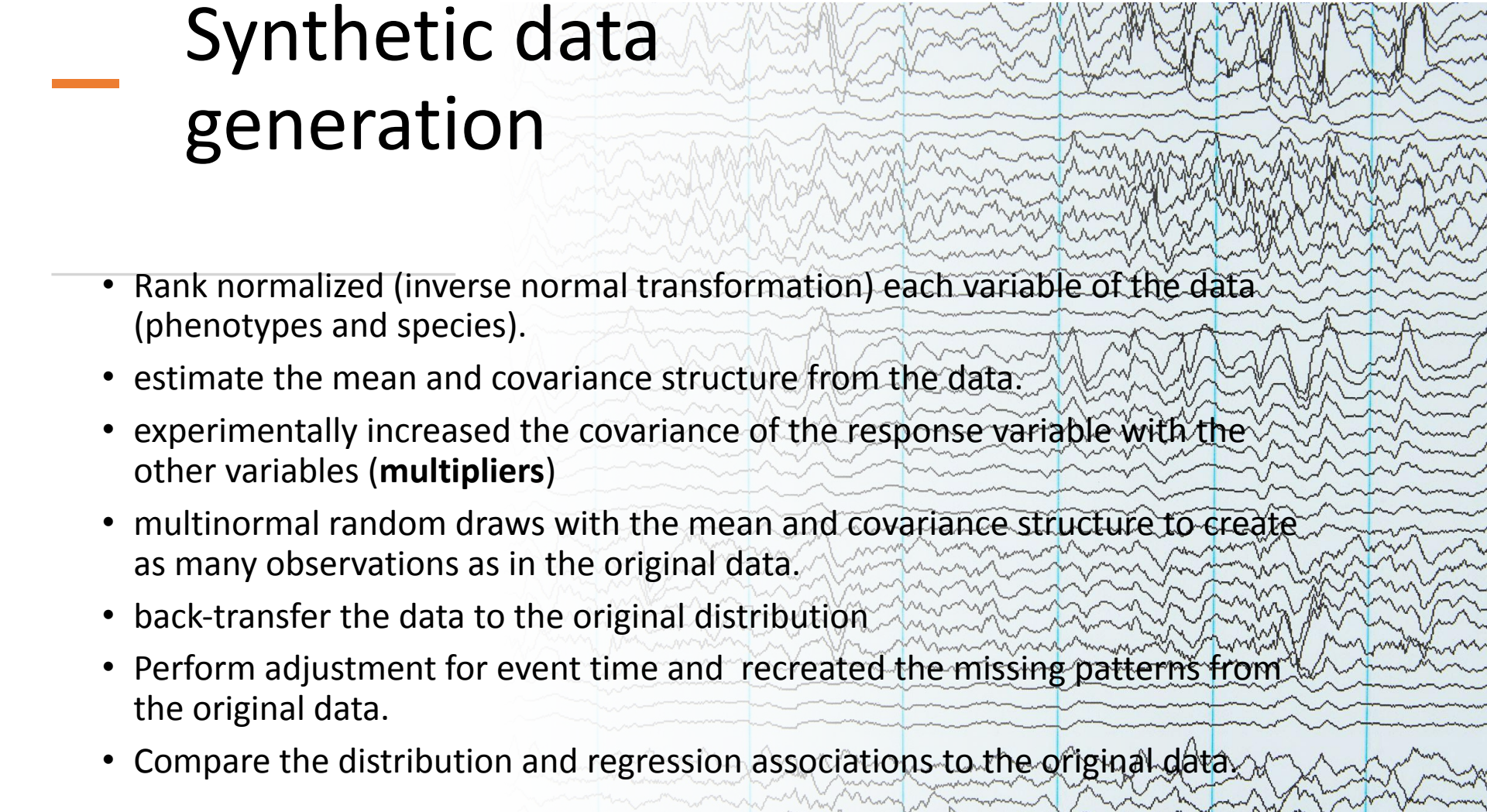
Number of invitee
randomly sampled from the National Population Information System

- Questionnaires (e.g socioeconomic position, use of health care service, past and current diagnose, diet and health behaviour)
- Health examination (e.g weight, height, blood pressure)
- Collections of biological samples (e.g blood, urine), **fecal samples were also collected in 2002**

Follow up through record linkage to national administrative registers

Borodulin et al., 2018

Synthetic data generation

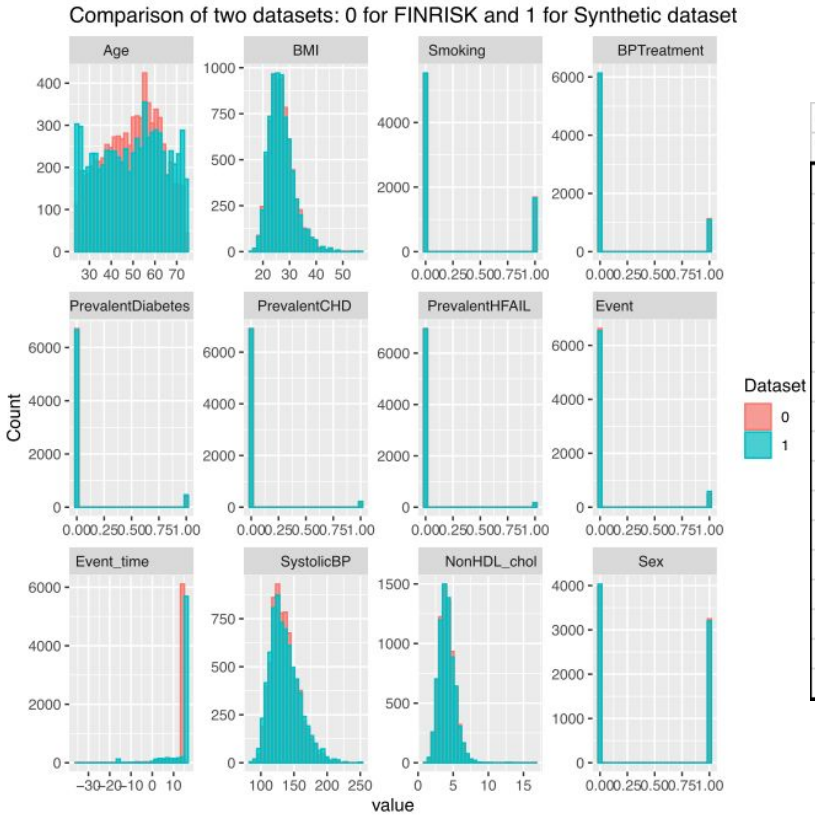


- Rank normalized (inverse normal transformation) each variable of the data (phenotypes and species).
- estimate the mean and covariance structure from the data.
- experimentally increased the covariance of the response variable with the other variables (**multipliers**)
- multinormal random draws with the mean and covariance structure to create as many observations as in the original data.
- back-transfer the data to the original distribution
- Perform adjustment for event time and recreated the missing patterns from the original data.
- Compare the distribution and regression associations to the original data.

Metadata covariates

ID	Label	Class	Levels
Sample_ID	Unique pseudonymized observation IDs		
Event	Incident heart failure status; excluding those occur before baseline	Integer	0=No;-1=Yes
Event_time	Time from baseline to the Heart Failure	Numeric	- 35.41* and + 16.9 years
Sex	Sex of individuals	Integer	0=Female; 1=Male
Age	Age of individuals at baseline	Numeric	24.1-74.24 years old
BodyMassIndex	kg/m2; at baseline	Numeric	15.84-56.94
Smoking	Smoking status at baseline	Integer	0=No;-1=Yes
BPTreatment	Hypertension treatment at baseline	Integer	0=No;-1=Yes
PrevalentDiabetes	Diabetes at baseline	Integer	0=No;-1=Yes
PrevalentCHD	Coronary heart disease at baseline	Integer	0=No;-1=Yes
SystolicBP	mmHg; systolic blood pressure at baseline	Numeric	88-253
NonHDLcholesterol	mmol/L; non-HDL cholesterol at baseline	Numeric	1.07-16.64
PrevalentHFAIL	Heart Failure diseases at the baseline	Integer	0=No;-1=Yes

Selected Synthetic dataset



Comparison of FINRISK vs Synthetic dataset				
Wilcoxon test for numerical data and Fisher test for categorical data				
	Variable	Synthetic Dataset	FINRISK Dataset	P-value
1	N	7231	7231	NA
2	Richness	195.22+-37.57(76.88-374.53)	189.46+-36.81(59.16-397.36)	1.1e-25
3	exp(Shannon)	32.2+-11.59(1.97-89.36)	31.37+-11.43(1.86-97.52)	2.8e-05
4	inv(Simpson)	15.1+-6.95(1.24-44.47)	14.83+-6.89(1.25-47.46)	0.017
5	evenness (q1)	0.16+-0.04(0.02-0.33)	0.16+-0.04(0.03-0.29)	0.24
6	Age	49.35+-14.83(24.1-74.24)	49.47+-12.97(24.1-74.24)	0.9
7	BodyMassIndex	26.97+-4.7(15.84-56.94)	27.01+-4.69(15.84-56.94)	0.43
8	SystolicBP	136.29+-22.09(87.9-253.07)	135.81+-20.33(88-253)	0.92
9	NonHDL_cholesterol	4.09+-1.1(1.07-12.99)	4.09+-1.09(1.07-16.64)	0.92
10	Event_time	13.76+-5.58(-23.71-16)	13.46+-4.27(-35.41-14.96)	0
11	Sex	3199 (44.2)	3248 (44.9)	0.42
12	Smoking	1663 (23)	1687 (23.3)	0.65
13	BPTreatment	1096 (15.2)	1129 (15.6)	0.46
14	PrevalentDiabetes	462 (6.4)	405 (5.6)	0.05
15	PrevalentCHD	221 (3.1)	205 (2.8)	0.46
16	PrevalentHFAIL	177 (2.4)	160 (2.2)	0.38
17	Event	582 (8)	493 (6.8)	0.0052

Model performance: Harrel's concordance index

- • A good measure for survival models when the data is censored
- • Patients with shorter times-to-event should have higher risk scores

Model Calibration: Hosmer-Lemeshow test at 15 years of follow-up

- • the estimated probabilities it outputs are accurate

Baseline Models

- Cox model with only Age + Sex covariates
- Cox model with all clinical covariates
- Cox model with all clinical covariates + microbiome data

Overall baseline model performance

	Real Dataset		Synthetic Dataset	
	Harrel's C	**Hosmer-Lemeshow**	**Harrel's C**	**Hosmer-Lemeshow**
Age+sex	0.815517316181654	0.4419724712633	0.723091773400037	9.539331163593e-99
All covariates	0.855461670628663	0.110166895232517	0.71097917920827	1.26281928124e-150
Microbiome+ all covariates	0.823621342999035	1.718778805022e-07	0.659159003314443	1.27643033558e-259

Leaderboard submissions phase

Submission ID	Date	Participant/Team	Harrells C	Hosmer-Lemeshow
9730535	1/10/2023 1:53 AM		0.964662114389025	0.0027391484623173
9730509	1/9/2023 3:51 PM		0.85550874385106	3.68376485511634e-165
9730516	1/9/2023 5:37 PM		0.85550874385106	3.6837648551155e-165
9730289	1/6/2023 6:03 PM		0.855499329206581	0.0000805580909573684
9730540	1/10/2023 3:46 AM		0.855461670628663	0.017109223667752
9730539	1/10/2023 3:40 AM		0.855461670628663	0.124062184456823
9730537	1/10/2023 2:52 AM		0.854619059947749	0.0003183735189758
9730494	1/9/2023 2:08 AM		0.853639936921882	1.43329135289717e-172
9729604	12/9/2022 10:58 AM		0.852905594652482	1.01380073567315e-302
9730252	1/3/2023 5:38 PM		0.852505472262104	0.626600030695395
9730529	1/9/2023 9:08 PM		0.851695812836868	0.176194374513841

Total accepted: 35 VALID submissions from 9 participants/teams

Leaderboard final round

id	createdOn	submitterid	status	harrell_c	hoslem_test
9731637	1/29/2023 11:29 PM	SB2	SCORED	**0.839395379197954**	0.0032924949185753
9731625	1/29/2023 7:12 PM	DenverFINRISKHacky	SCORED	**0.835136923343354**	0.0120891360271361
9731490	1/27/2023 1:26 AM		SCORED	**0.827974369335565**	2.40190686358651e-22
9731454	1/26/2023 11:55 AM		SCORED	**0.827752343482875**	7.48389464769304e-75
9731636	1/29/2023 11:02 PM		SCORED	**0.825842921149739**	1.14999761860238e-178
9731713	1/31/2023 12:33 AM		SCORED	**0.754830172425277**	2.18334641086665e-83
9731666	1/30/2023 9:35 PM		SCORED	**0.315609749599243**	0

More reading on the formula's idea

Bayesian perspective, and employ the Bayesian bootstrap to estimate the posterior distribution of the the statistic

$$\Delta s = s(\hat{\boldsymbol{y}}_t\,,\,\boldsymbol{y}) - s(\hat{\boldsymbol{y}}_{t-1}\,,\,\boldsymbol{y})$$

where s(yˆ , y) represents an arbitrary scoring metric, and perform a onesided Bayesian hypothesis test to determine whether the current score is statistically better than the best score so far. With the larger score the better, the one sided test is:

$$H_0 : \Delta s \leq 0 \quad \text{vs} \quad H_1 : \Delta s > 0$$

The Bayesian test is based on the posterior odds in favor of H_1,

$$PO = \frac{P(H_1 \mid D)}{P(H_0 \mid D)} = \frac{P(H_1 \mid D)}{1 - P(H_1 \mid D)},$$

Neto et. al, 2016

Opinion as organizer

- Level of participants exploiting the microbiome aspects varied
- The facts that participants failed to provide well-calibrated models despite of well score for accuracy raised a concern related to how well the machine learning methods could provide realistic models for predicting incident diseases
- Prior information of evaluations metrics and code can limit the willingness of participants to innovate
  - most of the motivations is to win

Participant point of view

- Work on synthetic data is challenging
- No comparable public datasets were identified, so leaderboard metrics had to be heavily relied on, without any feedback possible from visualizations or stdout
- Limited runtime cut-off also minimizes the chance for participants to improve the parameter for model
- Any other relevant metadata may support better interpretable model
- Due to limited access to real data, the hyperparameters of machine learning models can not be perfectly tuned.