AI i psykiatrien

Michael Eriksen Benrós
Professor, Chief Physician, Head of Research, MD, PhD
Copenhagen Research Centre for Biological and Precision Psychiatry,
Mental Health Centre Copenhagen, Copenhagen University Hospital
& University of Copenhagen

REGION H Mental Health Services

UNIVERSITY OF COPENHAGEN

PRECISE- PRECISION PSYCHIATRY INITIATIVE

DATA IS GOLD – bigger and better samples

- The most comprehensive **nationwide registers** (diagnoses, anti-infectious prescriptions, blood and CSF tests)

- **Electronic Health Records** population-based from ½ of Denmark
- The world's largest **genotyped cohort** with environmental data

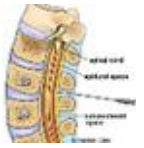
  iPSYCH: 135,000

  Copenhagen Hospital Biobank: 500.000

- The world's largest **CSF biobank** (>20,000 people) with **blood samples** from the same individuals (SSI)

- **DanFund population-based cohort** with phenotyping, blood, microbiome and genetics (N~10.000)

- **PSYCH-FLAME** clinical cohort with extensive psychopathology and biological measurements (CSF, blood, microbiome), with wearables, speech and facial recordings (N~500)



PRECISION PSYCHIATRY INITIATIVE (PRECISE)

Challenge:
Currently, choice of treatment for mental disorders is determined by trial and error using a “one-size-fits-all” approach resulting in an unacceptably large proportion of non-responding patients.

Limitations of prior research:
Focus on single exposures and single outcomes, not accounting for the complexity of mental disorders.

Solution:
Leveraging on the wealth of data and novel data analytical approaches now available. Developing more accurate predictive models integrating multiple objective measures would enable true precision psychiatry improving clinical decision making and treatment response.

Outcomes

Prediction models

- Diagnoses
- Acute re-admissions
- Suicide and attempts
- Treatment outcomes
- Coercion
- Adverse events
- Treatment response
- Patient trajectories
- Missing tests according to guidelines
- Who and when are brain scans indicated

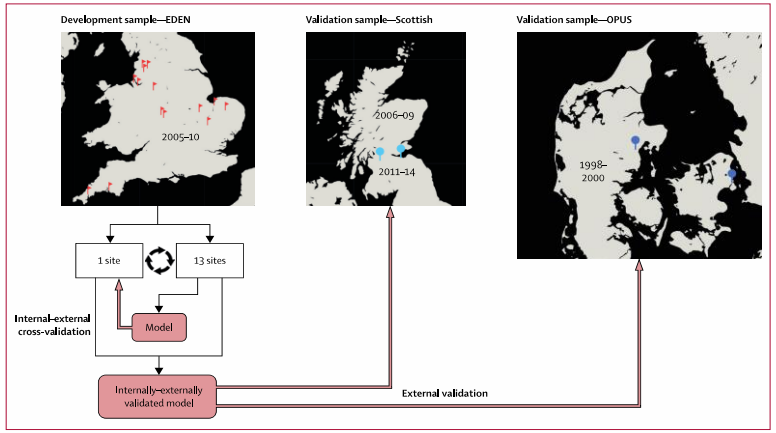
Novel insights

- Identify novel clinical markers and biomarkers associated
- Identify modifiable factors with a prevention potential
- Identify novel clusters within or across current diagnoses
- Obtain novel ground-breaking insights into mental disorders

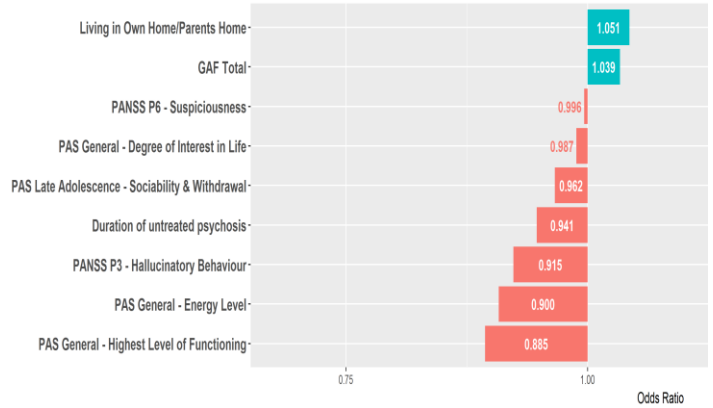
Innovations

- ***Decision support tools***
- Improved prevention and treatment of mental disorders
- Enhance the national AI competencies in the Region
- Facilitate the use of EHR data
- *Leaders internationally within precision psychiatry*

At First Episode Psychosis: predicting of A) Recovery, B) Social recovery, C) Vocational recovery, and D) Quality of life



International replication and validation



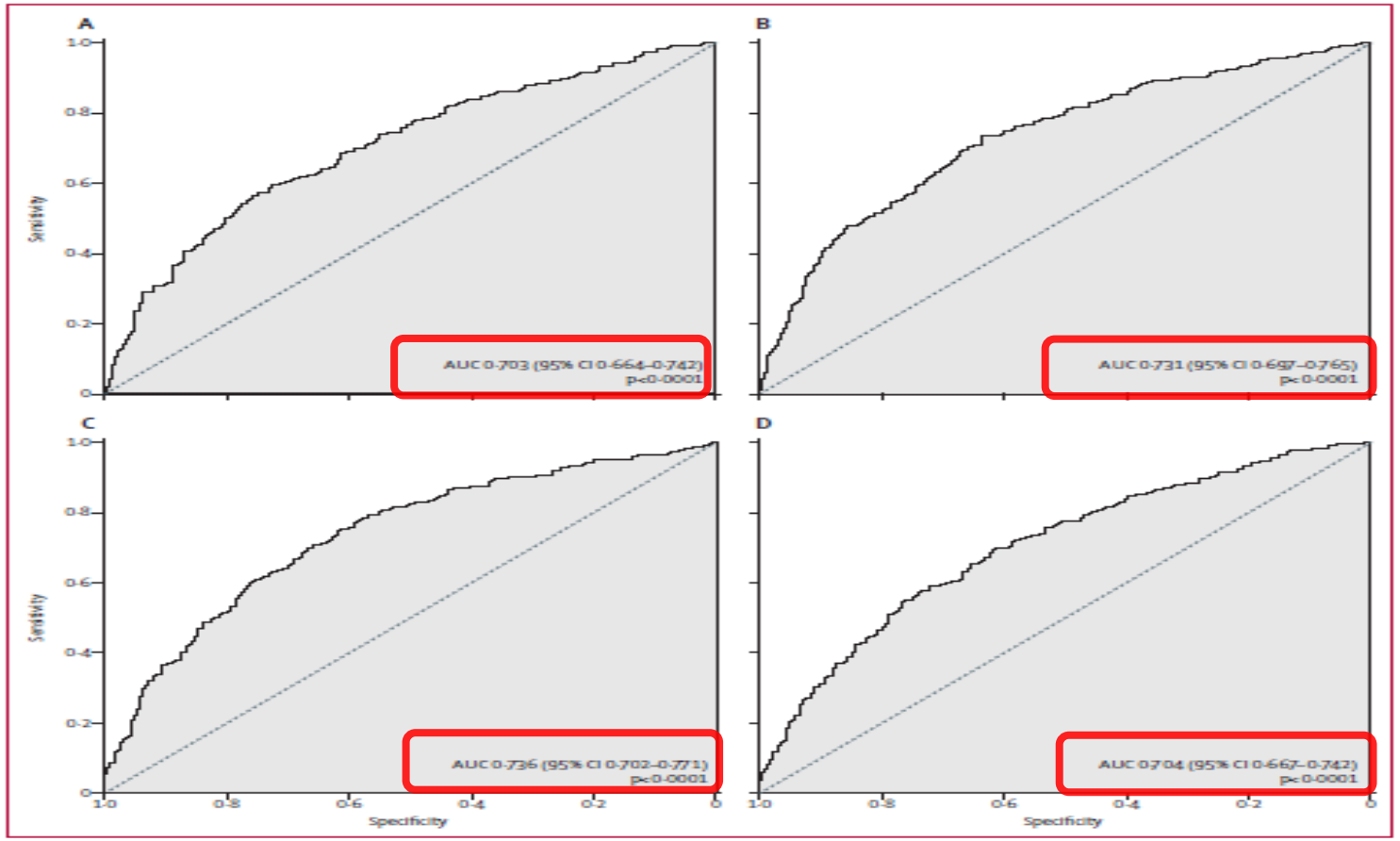
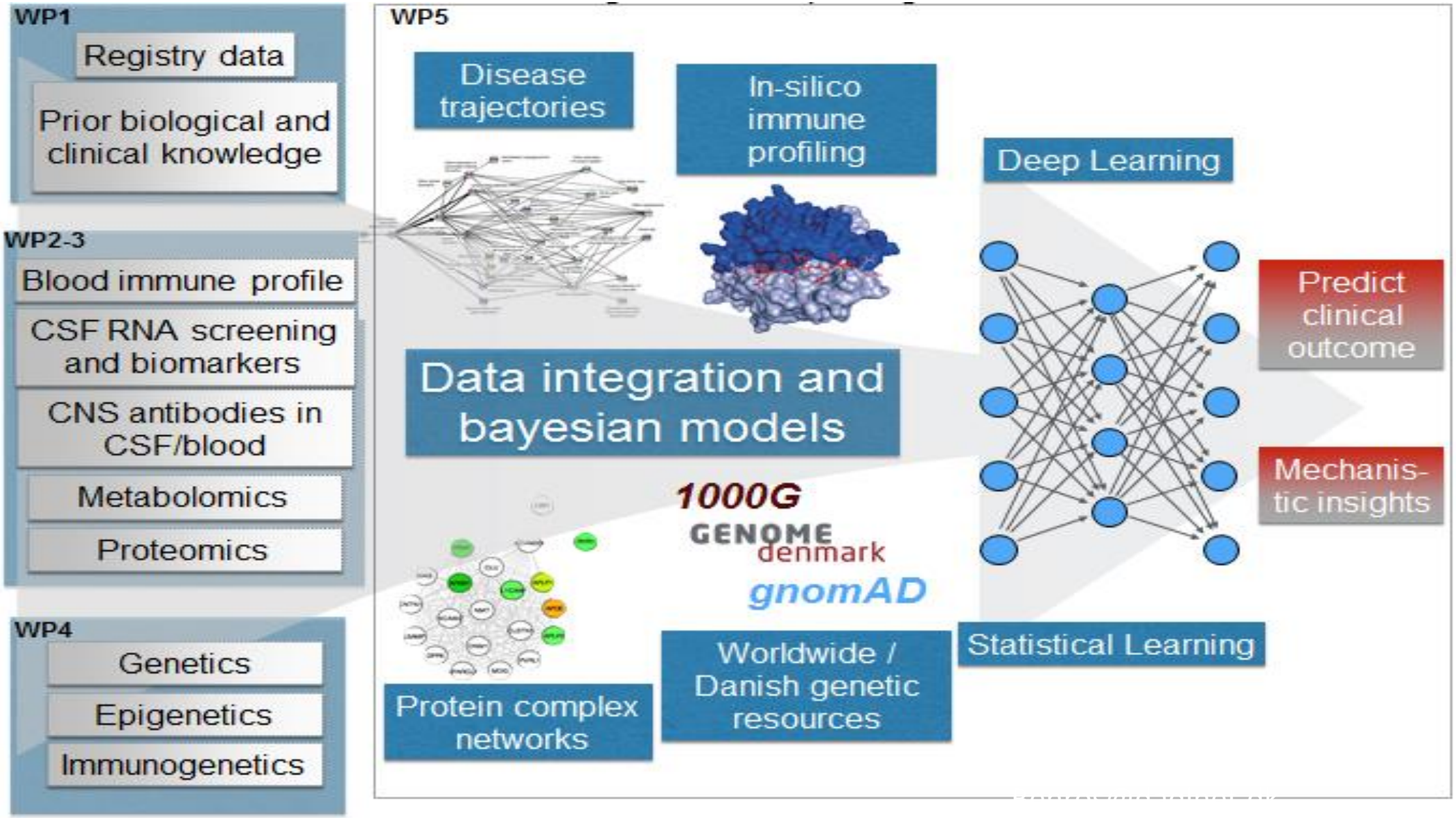


Figure 3: ROC curves showing internal–external LOSOCV model performance in the EDEN dataset for 1-year symptom recovery (A), social recovery (B), vocational recovery (C), and quality of life (D) models
ROC=receiver operating characteristic. LOSOCV=leave-one-site-out cross-validation. AUC=area under the curve.

Lancet Digital Health 2019;

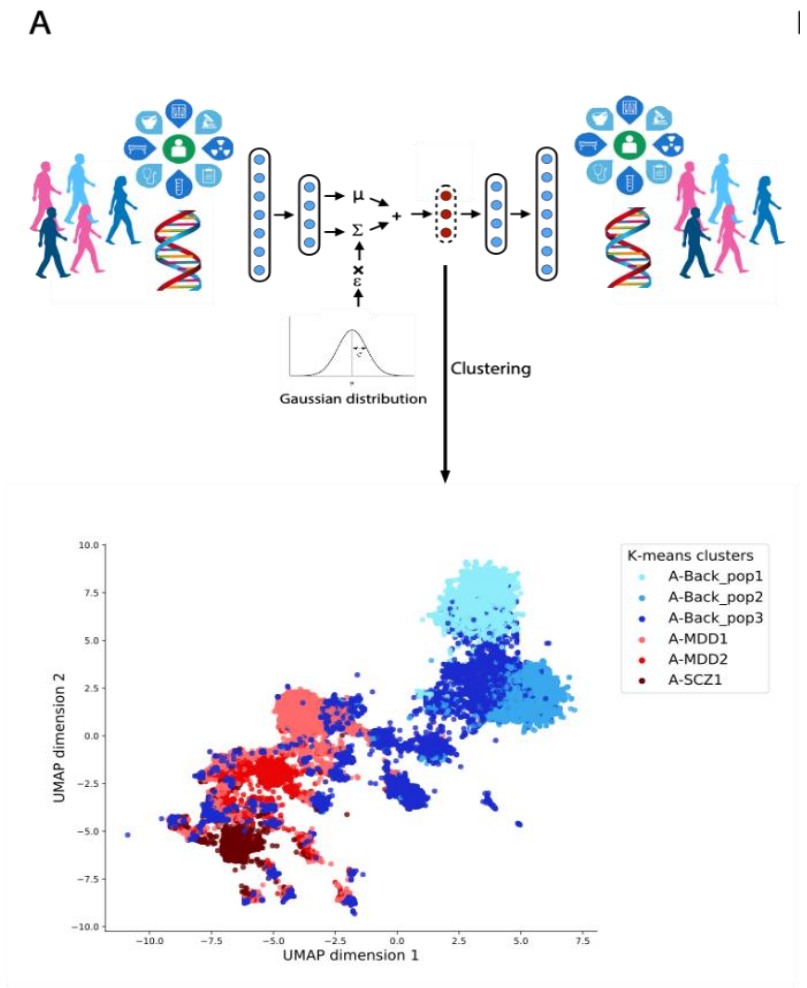
Predictors of poor outcome: Poor premorbid adjustment, Stable housing, Long DUP, Paranoia & Hallucinations,

Psych-Flame: Identifying immune subtypes

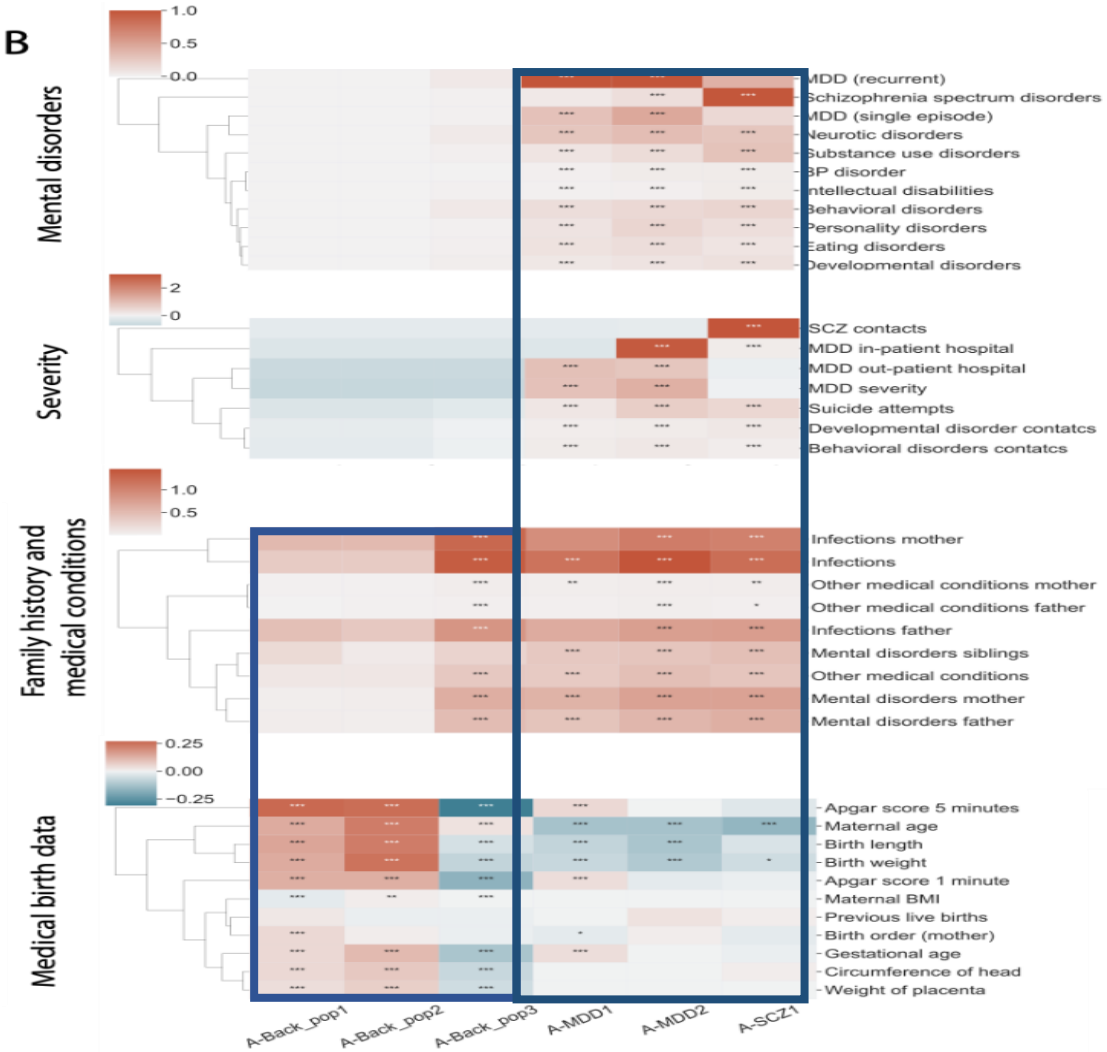


Clustering of schizophrenia and depression and HC (N=42,103)
Based on information prior to the diagnosis in the registers and genetics

Overview of the VAE framework for data integration



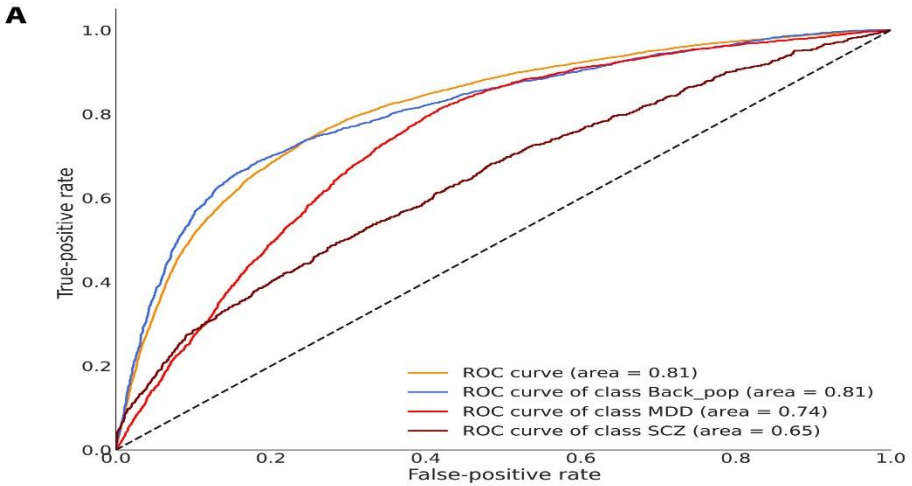
6 clusters overall according to severity



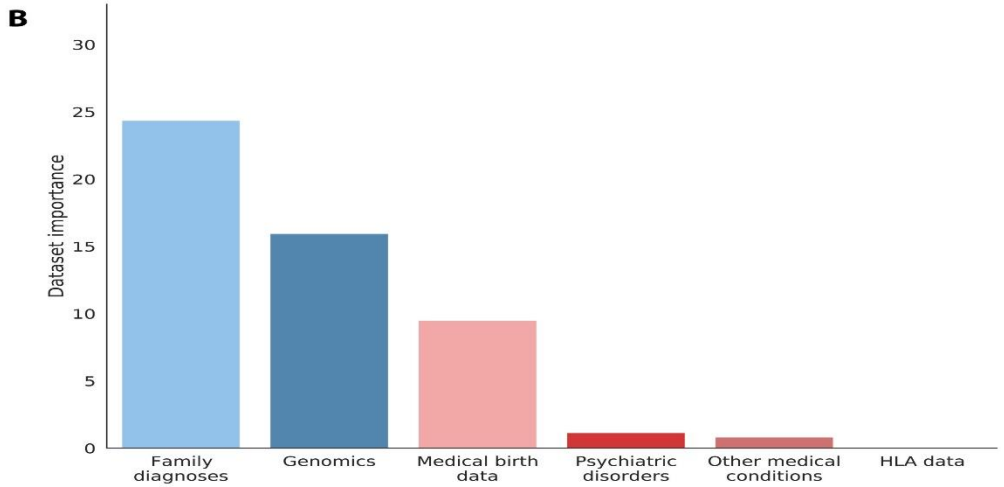
Allesøe RL… Rasmussen S, **Benros ME**. 2022. ***SCIENCE Adv***

Particularly family history, genetics and birth related factors are important to separate from the background population

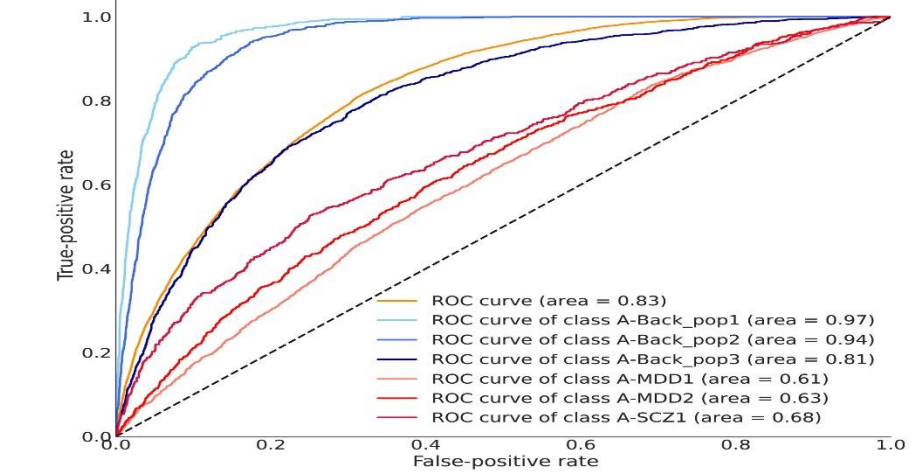
Diagnoses overall AUC=0.81

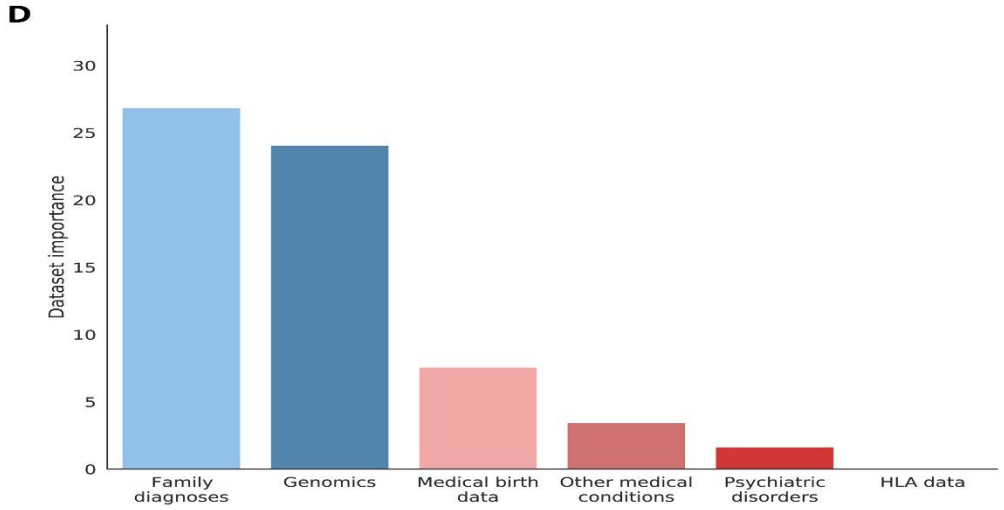


Variable importance grouped



C Clusters AUC=0.83

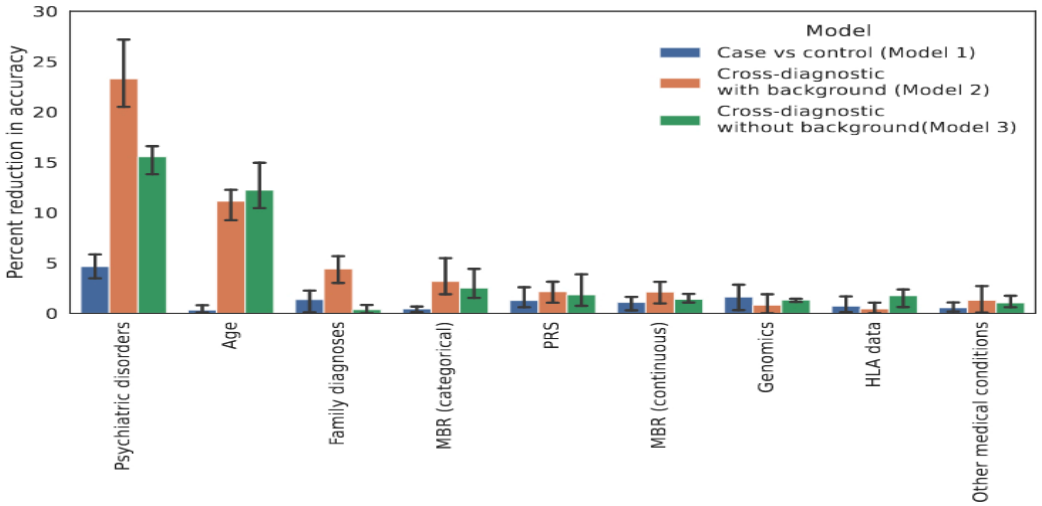




Allesøe RL… Rasmussen S, **Benros ME**. 2022. SCIENCE Adv

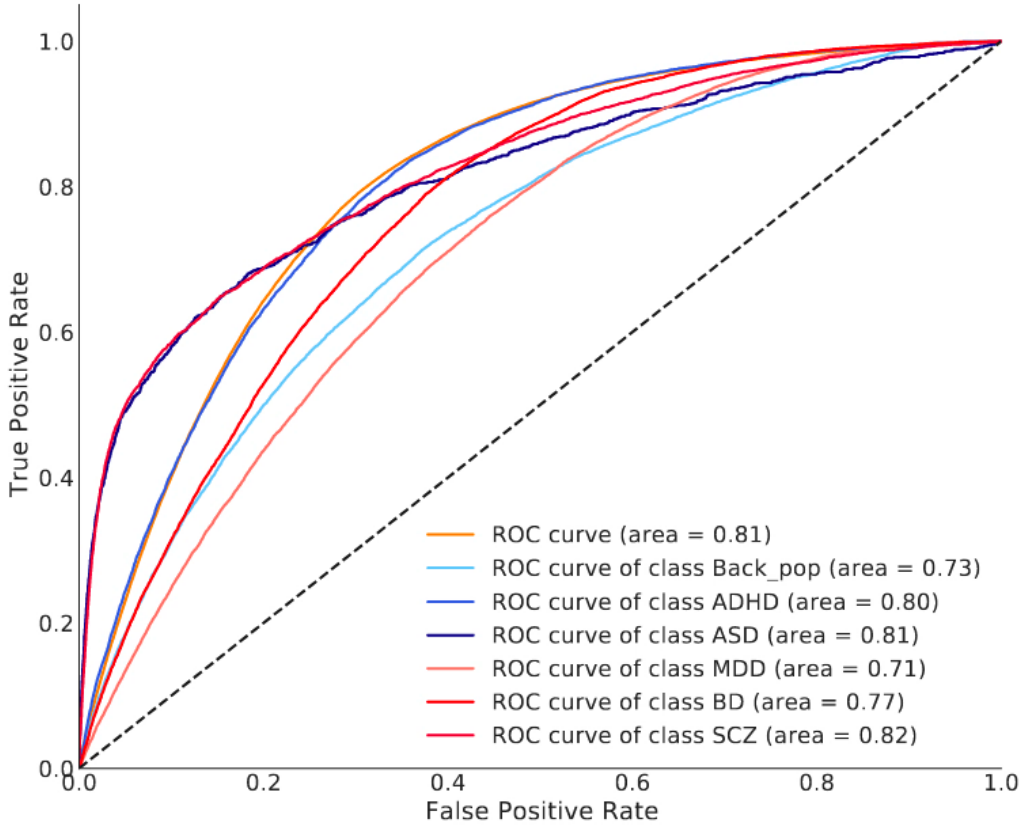
Deep Learning for cross-diagnostic prediction of mental disorder diagnosis and severity using nationwide registry and genetic data (N=78.445)

Variable importance



- 57,764 individuals with at least one of the following mental disorders:
  - 15,969 with attention-deficit/hyperactivity disorder (**ADHD**)
  - 12,878 with autism (**ASD**)
  - 19,159 with major depressiv disorder (**MDD**)
  - 5120 with schizophrenia or psychotic episodes (**SCZ**)
  - 1,719 with bipolar disorder (**BD**)
- 20,681 background population group (**Back_pop**)

Model AUC of 0.81



Allesøe RL... Rasmussen S, **Benros ME**. 2022. **JAMA Psychiatry**

Performance Measures in the Multiclass Prediction Model

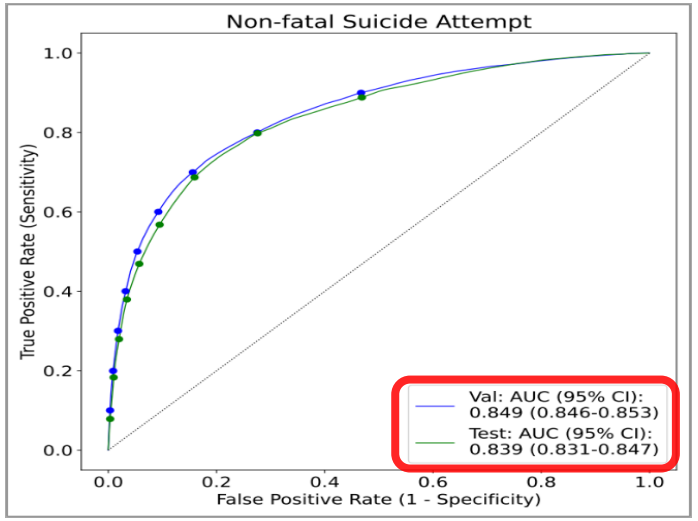
Class	Accuracy	Precision	Sensitivity	Specificity	AUC
Model 2 (cross-diagnostic prediction including population control group)					
Back_pop	0.68	0.52	0.48	0.78	0.73
ADHD	0.80	0.35	0.34	0.89	0.80
ASD	0.80	0.42	0.50	0.86	0.81
MDD	0.75	0.50	0.47	0.84	0.71
BD	0.96	0.18	0.26	0.98	0.77
SCZ	0.90	0.40	0.41	0.95	0.82
Model 3 (cross-diagnostic prediction of cases only)					
ADHD	0.74	0.42	0.39	0.84	0.73
ASD	0.78	0.56	0.58	0.84	0.83
MDD	0.75	0.66	0.67	0.79	0.80
BD	0.94	0.18	0.24	0.97	0.81
SCZ	0.87	0.44	0.40	0.93	0.80

Abbreviations: ADHD, attention-deficit/hyperactivity disorder; ASD, autism spectrum disorder; AUC, area under the curve; Back_pop, background population control; BD, bipolar disorder; FN, false negative; FP, false positive; MDD, major depressive disorder; SCZ, schizophrenia spectrum disorders; TN, true negative; TP, true positive.

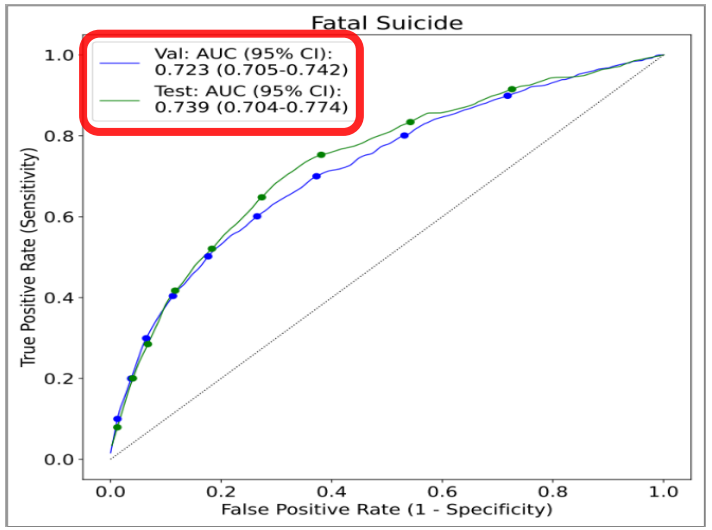
[a] Performance as accuracy, precision (positive predictive value), sensitivity (recall), specificity, and AUC for each diagnostic category in the multiclass models. The evaluations are done by considering each class/diagnosis separately in the model and collapsing all other classes into one. In the table, accuracy is calculated as TP + TN / TP + TN + FP + FN; precision as TP / TP + FP; sensitivity as TP / TP + FN; and specificity as TN / TN + FP.

Register prediction of suicidal attempts and suicide

Predicting suicidal behavior



Predicting suicide



Performance matrix
Not ready for clinical use

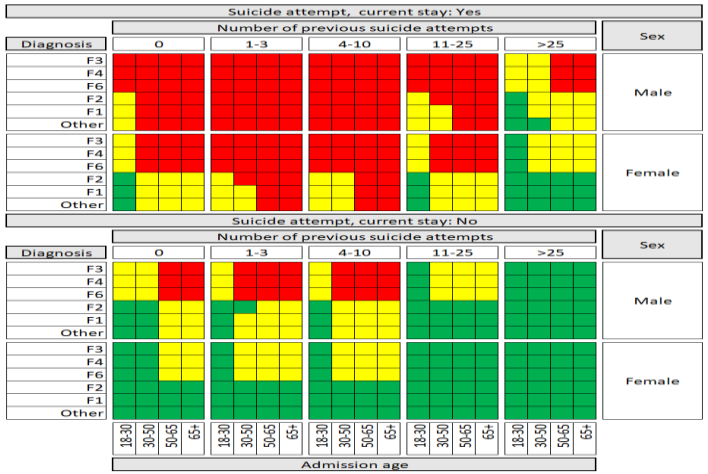
Non-fatal suicide attempt

sensitivity	sensitivity	specificity	PPV	NPV	threshold	TP	TN	FP	FN
10%	0,1000	0,9968	0,3213	0,9865	0,2473	1093	718088	2309	9834
20%	0,1999	0,9910	0,2524	0,9879	0,1557	2184	713929	6468	8743
30%	0,3000	0,9821	0,2027	0,9893	0,1035	3278	707504	12893	7649
40%	0,4000	0,9680	0,1594	0,9907	0,0686	4371	697351	23046	6556
50%	0,5000	0,9462	0,1235	0,9920	0,0451	5464	681625	38772	5463
60%	0,6000	0,9076	0,0897	0,9934	0,0270	6556	653866	66531	4371
70%	0,7000	0,8440	0,0637	0,9946	0,0162	7649	608026	112371	3278
80%	0,8000	0,7247	0,0422	0,9958	0,0096	8742	522072	198325	2185
90%	0,9000	0,5330	0,0284	0,9972	0,0062	9834	383973	336424	1093

Fatal sucide

sensitivity	sensitivity	specificity	PPV	NPV	threshold	TP	TN	FP	FN
10%	0,1004	0,9877	0,0093	0,9990	0,0048	84	721501	8986	753
20%	0,1995	0,9628	0,0061	0,9990	0,0032	167	703304	27183	670
30%	0,2999	0,9353	0,0053	0,9991	0,0024	251	683259	47228	586
40%	0,4002	0,8866	0,0040	0,9992	0,0019	335	647616	82871	502
50%	0,5006	0,8233	0,0032	0,9993	0,0015	419	601405	129082	418
60%	0,5998	0,7346	0,0026	0,9994	0,0012	502	536602	193885	335
70%	0,7001	0,6274	0,0021	0,9995	0,0010	586	458302	272185	251
80%	0,8005	0,4681	0,0017	0,9995	0,0008	670	341929	388558	167
90%	0,8996	0,2815	0,0014	0,9996	0,0007	753	205650	524837	84

Most important variables for the model



Nielsen SD… Clemmensen L, **Benros ME**. 2023. **Acta Psychiatry Scandinavia**

Register prediction study of psychiatric re-admissions

Acute Psychiatric Re-admission Risk at Discharge (APRAD)

APRAD

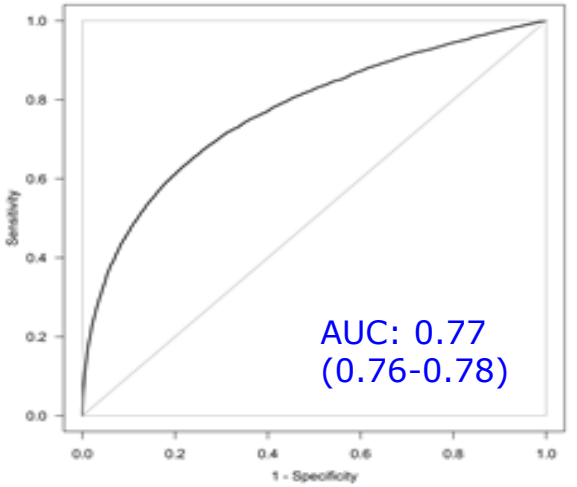
Number of Acute Psychiatric Re-admissions last 12 months

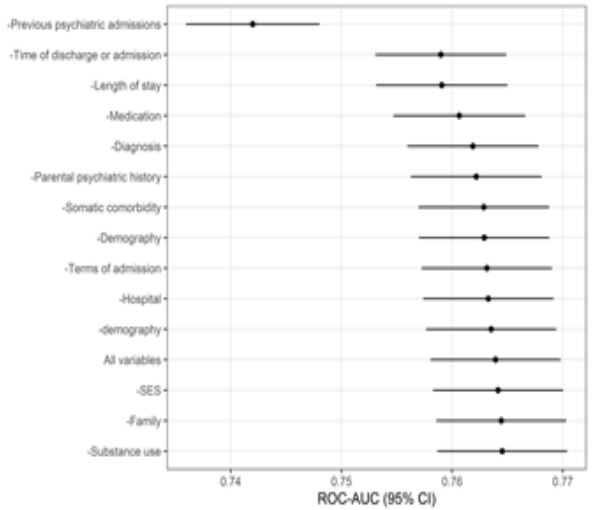
Diagnosis	0		1		2-5		>5	
	<4	1-4	<4	1-4	<4	1-4	<4	1-4
Disorder of personality and behavior (F60-69)	16	14	24	28	41	52	71	79
Schizophrenia and related disorders (F20-29)	14	20	24	34	40	47	68	77
Affective disorders (F30-39)	11	16	16	38	26	53	64	84
Disorder due to substance use (F10-19)	16	14	25	28	32	46	63	66
Other mental disorder diagnosis	13	13	22	26	32	49	60	70

Length of hospital stay (days)

Absolute risk of acute psychiatric re-admission

10-14% | 15-24% | 25-34% | 35-44% | 45-64% | ≥65%





Bach-Mortensen et al. In review

SP prædiktionsmodel af akutte genindlæggelser

Model	Accuracy	Sensitivity	PPV	Specificity	NPV	F1-score	MCC	AUC
1F-model unstable pathway	79.1%	40.7%	73.9%	94.3%	80.0%	0.525	0.435	0.68
1F-model previous acute readmissions	77.3%	52.2%	61.8%	87.2%	82.1%	0.566	0.416	0.74
2F-model Unstable + prev acute	80.0%	58.0%	67.0%	88.7%	84.2%	0.622	0.489	0.78
3F-model register's*	76.6%	55.2%	59.4%	85.0%	82.7%	0.573	0.412	0.76
5F-model register's*	75.9%	57.5%	57.5%	83.2%	83.2%	0.575	0.407	0.77
5F-model + unstable pathway	79.8%	58.0%	66.5%	88.4%	84.1%	0.620	0.485	0.79
Best model – 9th May	79.3%	62.0%	64.1%	86.2%	85.1%	0.630	0.487	0.80
Best model - 25th May *(preliminary results)*	**80.0%**	**73.9%**	**78.5%**	**84.6%**	**81.0%**	**0.761**	**0.590**	**0.85**

Almost **75%** of them will be detected, and **8 out of 10** predicted acute readmissions are correct

85% non-acute discharges will be detected and **8 out of 10** times that a non-acute is predicted, it will be correct

Detection and security of acute readmissions as well as overall model performance is quite good

* - ***3F-model:*** *Length stay, primary diagnosis, number previous acute readmissions (1 year)*
- ***5F-model:*** *3F-model + Substance use diagnosis + F2 diagnosis*

Timeline & choosing wards/departments (acute re-admissions)

Choosing wards for phase 1:
- Stable leadership pre/during/post testing
- Stable acute re-admission rates
- Ward that would like to work with the project

Suggestions
- 1 inpatient ward
- 1 inpatient and 1 out patient team
- **1 connected in- and outpatient team (BBUI) + independent 1 inpatient ward + 1 outpatient ward**

Timeline:
- ***July/august****: error finding in the model script and data loading when automatized*
- ***September****: RHP quality and data team helps with variable and definition checks, and with their inputs from prior work including list of potential interventions*
- ***End October****: final model 1.0 ready for testing and made as Python package ready to load*
- ***October/November****: Decide if model is ready for clinical testing*
- ***Nov-Dec:*** Implement model in SP (informed by SP teams' prior experience and learnings)
- ***January****: SP module ready for phase 1 testing*
- ***Jan-March:*** *phase 1 testing*
- ***March 2024:*** *Evaluation of testing and if ready to proceed to phase 2 larger testing and potentially with further improved model based on the clinical feedback and with add-on NLP approach*

Clinical testing and implementation – new RHP methodology?

<u>Phase 1</u>: small scale clinical testing
- 1-3 clinical wards (dedicated to using it)
- Feasibility & acceptability
- **Can we make it work in clinical practice?**
- Do we see indications of results? (visual inspection of acute readmission graph)

<u>Phase 1</u>: Pilot
- Typisk på et afsnit.
- Arbejdsgang værktøj udvikles og resultater følges med ekstra stor støtte.
- Kan vi se resultater og forstår vi hvorfor?
- Evaluering og beslutning om at fortsætte

<u>Phase 2</u>: medium scale clinical testing
- ***Preliminary indications of efficacy*** of model in clinical use (non dedicated users)
- Further evaluate potential side effects
- Calculate needed sample size for phase 3 based on the effect estimates

<u>Phase 2</u>: Forberedelse til bred implementering
- Kan vi få det til at virke andre steder?
- Udvikling af metode og værktøjer, samt standardiseret implemeneteringsplan til bred implementering
- Implementering på 3-5 afsnit
- Evaluering og beslutning om at fortsætte

Clinical testing and implementation – new RHP methodology?

Phase 3: Efficacy trial
- Predefined power calculation based on phase 1-2 and minimum N and time
- As compared to TAU
- Randomized or block-randomization
- (But when more models are together it should be placebo/active controlled)

Phase 3: Bred implementering
- Udrulning integreres i hospitalets ledelseskæde og der sættes fokusmål for implementering.
- Fremdrift følges af HL og CL.

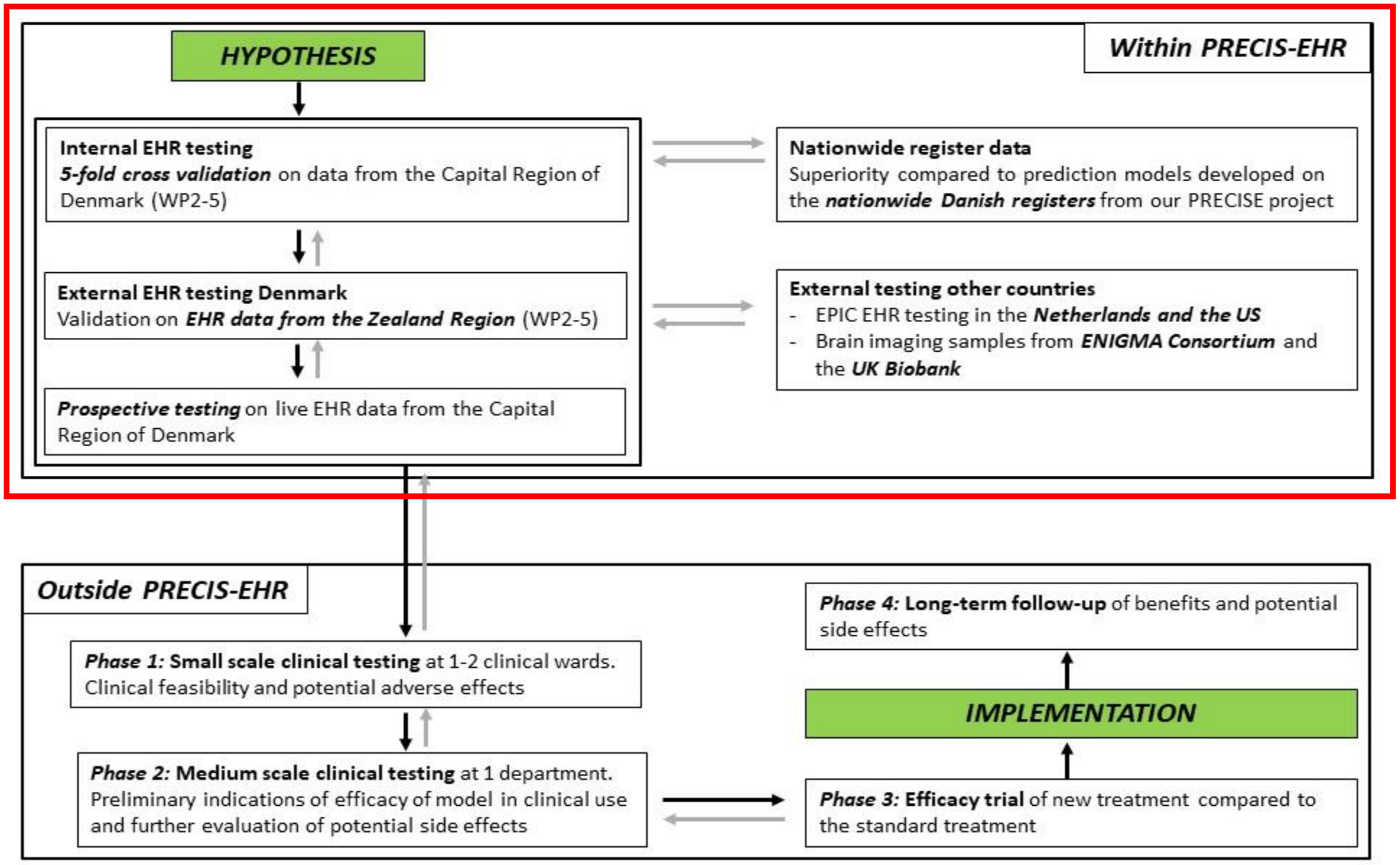
IMPLEMENTATION

IMPLEMENTATION

Phase 4: long-term follow-up
- Post implementation evaluation of benefits and potential side effects
- Tracking model performance and likely re-training/finetuning of model

Phase 4: Fastholdelse af resultater
- Opnåede resultater følges som hospitalsmål.
- Korrigerende handlinger træffes ved behov.

Testing and validation framework



Digital phenotyping of Mental Disorders

Quantifying the subjective psychiatric symptoms for better treatment

Challenge: Currently, there are no truly objective measurement in the psychiatric examination of the spoken words, facial expression and behavior.

Solution: Conduct **novel digital phenotyping** of mental disorders with AI

Part of Precision Psychiatry Initiative (PRECISE): Voice and speech analytics, facial analytics and wearables to predict psychiatric diagnosis and outcomes – in progress on the PSYCH-FLAME study

Pilot study: For the first time **use large-scale audio data** linked with EHR available from **Copenhagen Emergency Medical Services calls** (more than 5 million stored calls linked with EHR)

Perspective: Decision support tools assisting the health care professional while having the dialogue with the patient

Clinical AI Test Center

Big data derived prediction models

New risk factor candidates

Modifable Risk and Protective Factors

Prioritisation of Risk Factors and Prediction Models

Validated Prediction Models

Feedback from patients and clinicians

Clinical Test and Knowledge Center

B1. Understanding Patient and Clinician Needs for Data-Driven Treatments

- Panel of key stakeholders and end-users
- Qualitative investigations among focus groups
- Large-scale questionnaire investigations
- Investigate legislation and regulatory barriers
- Include novel discoveries

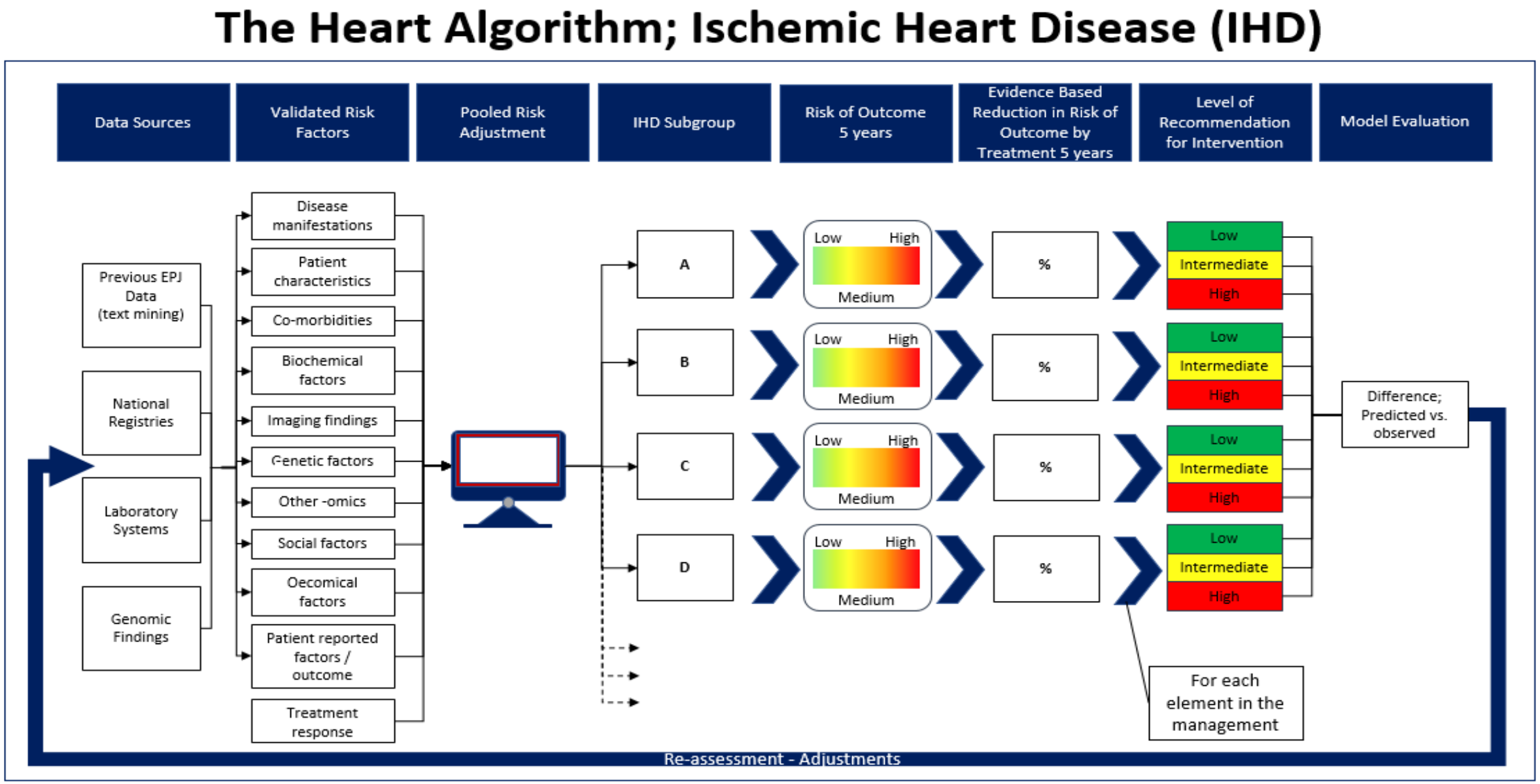
B2. Creating a Knowledge Center for Effective Data-driven Treatment and prediction models

B3. Clinical Test Center

- “laboratory” evaluating feasibility, acceptability and efficacy of secured webportal for individual risk predictions that can then be tested and implemented on EHR

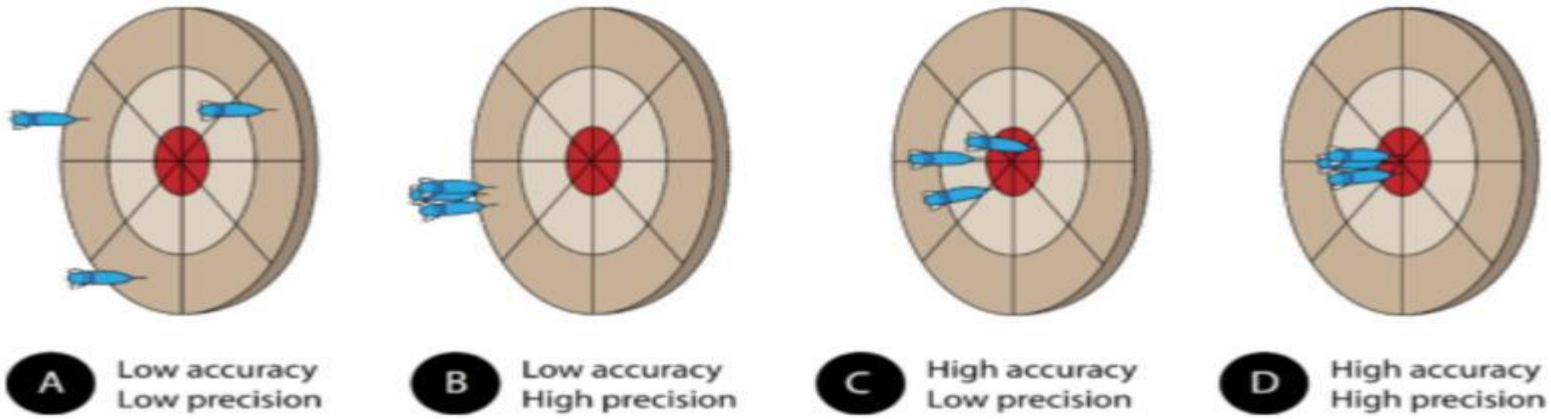
Towards Clinical Impact of EpiPsych Findings

Computer assisted risk assessment for precision medicine



Not only prediction…

- We need to **stratify patients based on meaningful variables** and ideally causal factors to achieve reproducible high accuracy and precision justifying implementation
- **Cross-disciplinary approach that includes:** mental health experts, bioinformaticians, engineers, data science and input from the end-users etc.
- Implementation should first occur after successful clinical testing



IMPACT

Paradigmatic shift

- Increase our ***understanding*** of mental disorders and the ***predictability*** of outcomes
- Identify causal factors for mental disorders
- Developing data driven ***decision support tools*** for mental disorders

Translational impact

- Identification of ***novel risk factors and clinical (bio)markers*** associated with mental disorders
- Paving the way for ***new treatment principles***
- ***Improved diagnostics and treatment***
- Striving towards ***implementation***

Vision:

- Objective biomarker-based treatment and data driven decision support based on a multitude of objective clinical markers for more personalized and better treatment

Thank you for your attention!

Thanks to my collaboraters:

Mental Health Center Copenhagen, Copenhagen University Hospital:

Rune Christensen, Rosa Allesøe, Enric Coppulo, Sara Nielsen, Terne Jakobsen, Jonas Meisner, Merete Nordentoft, Trine Madsen,

Human Protein Variation Group, University of Copenhagen:

Simon Rasmussen

Danish Technical University (DTU):

Line Clemmensen

Centre for Register-based Research, Aarhus University:

Preben Mortensen, Liselotte Petersen, Marianne Pedersen, Carsten Pedersen

Institute of Biological Psychiatry, Copenhagen University Hospital

Wes Thompson, Thomas Werge, Morten Krebs, Yunpeng Wang

Copenhagen Emergency Medical Services:

Helle Christensen, Frederik Folke

@MichaelBenros

Benros@dadlnet.dk