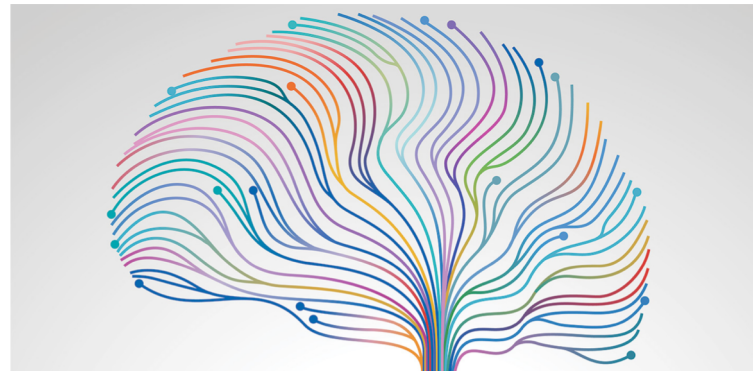




**University of
Zurich**^{UZH}



Prediction of Psychiatric Multimorbidity in a Large Pediatric Sample



Prof. Nicolas Langer
University of Zurich

Department of Psychology
Methods of Plasticity Research

Dr. Stefan Haufe
Charité

Universitätsmedizin Berlin
Berlin Center for Advanced Neuroimaging (BCAN)



4rd meeting of FG-AI4H
Shanghai World Expo Center China,
April 3rd 2019

Update: continuation of data collection (currently ~1800 subjects)

Data availability: Sample

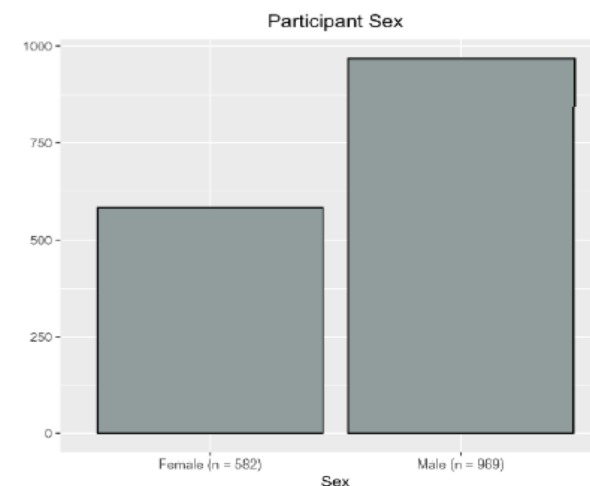
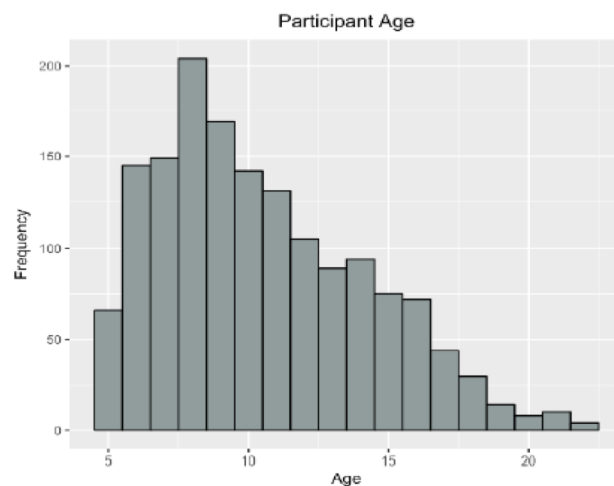
Healthy Brain Network (HBN) sample

Training Data:

- current release: 1602 subjects
- Age 5-21 years
- Population: typical developing children and children with psychiatric developmental disorders (~70/ multimorbidities)

Test Data:

- Subsample of training data
- Future release: approx. 500 subjects / year



Data availability: Sample

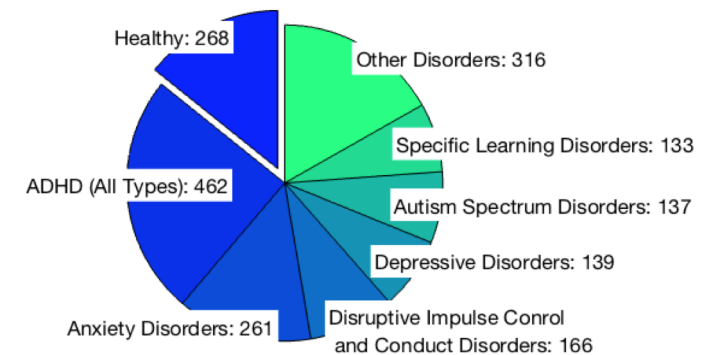
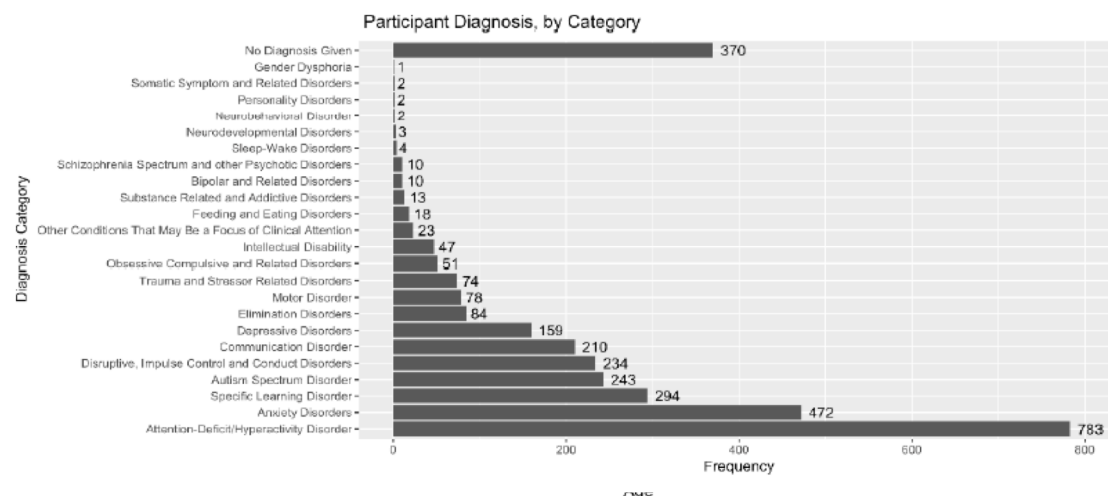
Healthy Brain Network (HBN) sample

Training Data:

- current release: 1602 subjects
- Age 5-21 years
- Population: typical developing children and children with psychiatric developmental disorders (~70% multimorbidities)

Test Data:

- Subsample of training data
- Future release: approx. 500 subjects / year



Data availability: Data

- Demographics
 - Age, gender
- Cognitive Data
 - e.g. WISC
- Behavioral Data
 - Questionnaires (SWAN)
- resting EEG
 - Raw data
 - Preprocessed data
 - EEG features
 - e.g. theta-beta ratio, alpha asymmetry
- Possibly T1-weighted MRI images
 - Source reconstruction
 - Cortical thickness



- **Prediction of Diagnosis**
 - DSM-V consensus diagnosis
- **Annotation Quality:**
 - based on the decision of a clinical team
 - all interviews and materials conducted as basis for the DSM-5 consensus diagnosis
 - conducted by licensed clinicians

Data availability: Data

- Demographics
 - Age, gender
- Cognitive Data
 - e.g. WISC
- Behavioral Data
 - Questionnaires (SWAN)
- resting EEG
 - Raw data
 - Preprocessed data
 - EEG features
 - e.g. theta-beta ratio, alpha asymmetry
- Possibly T1-weighted MRI images
 - Source reconstruction
 - Cortical thickness

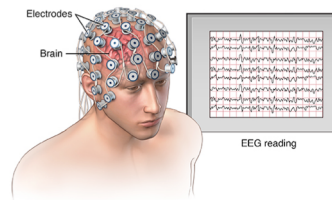
Cognitive & Behavioral Data:

- Demographics
- Cognition / Intelligence (e.g. WIAT, WISC-V, NIH-Toolbox)
- Medical history (e.g. addiction family history)
- Family structure, stress and trauma (negative life events, parenting)
- Personality traits (Big 5, self-esteem)
- Coping Strategies (communication skills, interpersonal factors)
- Physical measures (e.g. bio-electric impedance analysis, BMI, Metabolic rate, heart rate, blood pressure, height, weight, handedness,...)
- Social status (SES, parents education, family structure)

Nr. of features: ~270 (self-/ parent-/ teacher-report)

Data availability: Data

- Demographics
 - Age, gender
- Cognitive Data
 - e.g. WISC
- Behavioral Data
 - Questionnaires (SWAN)
- **resting EEG**
 - Raw data
 - Preprocessed data
 - EEG features
 - e.g. theta-beta ratio, alpha asymmetry
- Possibly T1-weighted MRI images
 - Source reconstruction
 - Cortical thickness



Raw EEG:

- 5 min.
- Eyes closed (40 s) & eye open (20 s)
- 128 electrodes (Geodesic EGI system)
- sampling rate 500 Hz
- **Nr. of features: ~ 150'000**



Update:

- new release of stable Automagic version (2.3.5), submitted paper revision
- Currently preprocessing all data

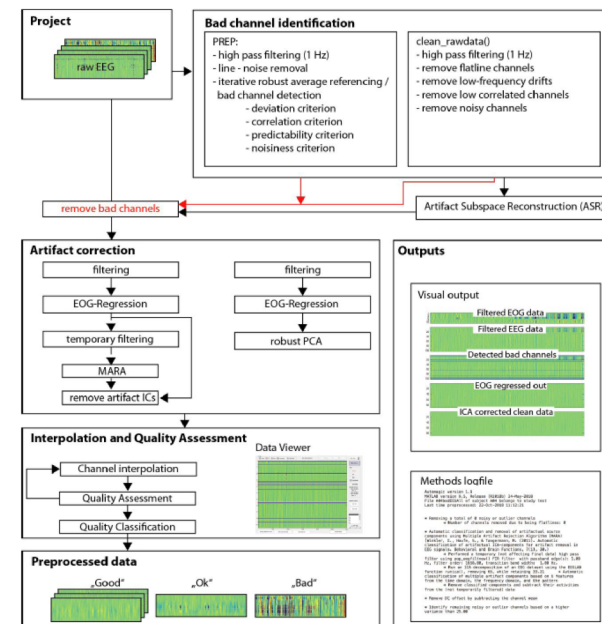
Prerequisite for Biomarker Research: Reliability of measures

Prerequisite for Reliability: Standardized Preprocessing

- Demographics
 - Age, gender
- Cognitive Data
 - e.g. WISC
- Behavioral Data
 - Questionnaires (SWAN)
- resting EEG
 - Raw data
 - Preprocessed data
 - EEG features
 - e.g. theta-beta ratio, alpha asymmetry

Automagic

Pedroni, Bahreini Langer, (2018), biorXiv



<https://github.com/methlabUZH/automagic>

Preprocessed EEG:

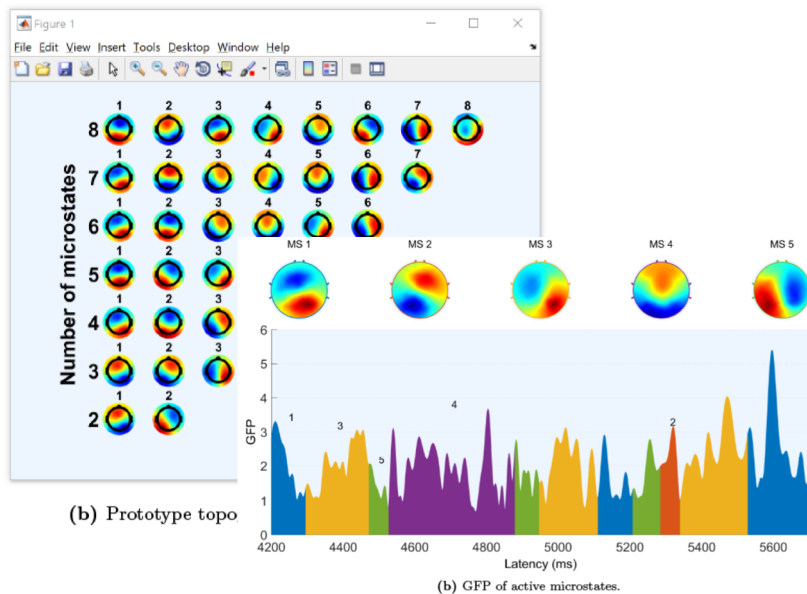
- Number of features: ~ 150'000

Update:

- Working on pipeline for functional connectivity features

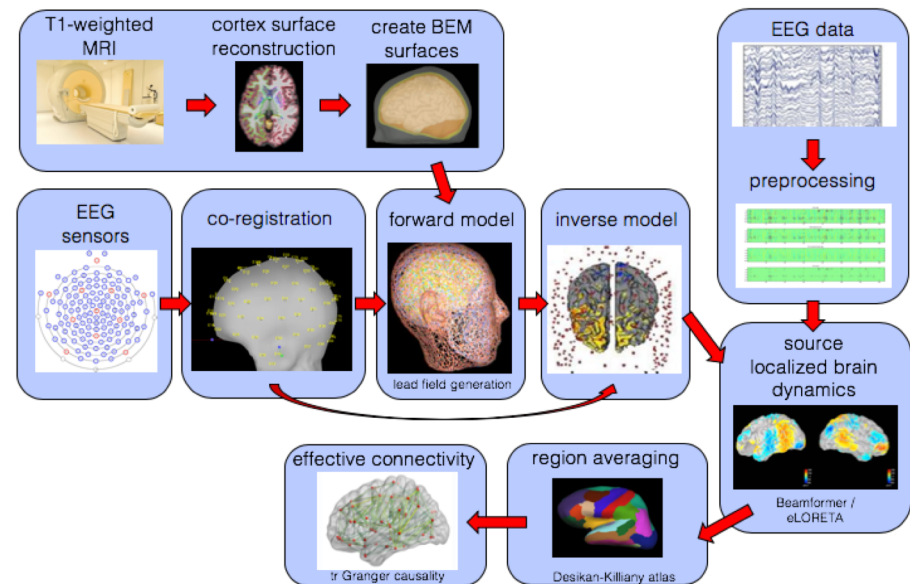
Developing Methods for EEG analysis

EEG Microstates Toolbox



Poulsen, Pedroni, Langer, Hansen (2018)

EEG Connectivity Analysis



Haufe & Langer in prep.

Update:

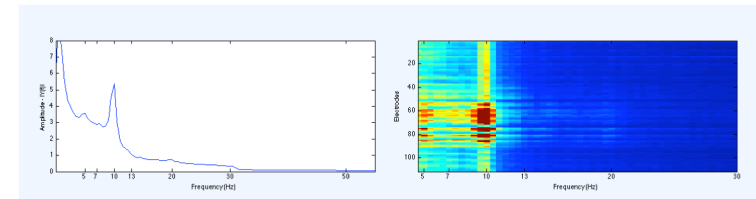
- Extracting frequency and time domain features

EEG features

- Demographics
 - Age, gender
- Cognitive Data
 - e.g. WISC
- Behavioral Data
 - Questionnaires (SWAN)
- **resting EEG**
 - Raw data
 - Preprocessed data
 - EEG features
 - e.g. theta-beta ratio, alpha asymmetry

Frequency Domain:

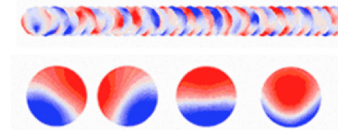
- Frequency Power analysis
 - (e.g. theta/beta ratio; alpha assymetry; 1/f noise, alpha peak)
- **Number of features: ~ 122**



Time Domain:

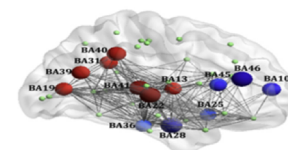
- Microstates:
 - „MS are stable spatial configurations of the electric field. These spatially stationary microstates might be the basic building blocks of information processing.“ (Lehmann, 1978)

- **Number of features: ~ 40**



Functional Connectivity:

- Imaginary part of coherency
- Time-reversed Granger causality
- **Number of features: ~ 9216**



Benchmarking

Task: prediction of multiple disorders from demographic, phenotypical (cognitive and behavioral) and EEG data

Training: on public HBN data

Benchmarking: on future releases of HBN data sets (approx. 500 subjects / year)

Implementation: participants submit executable code

- Standardized input (data folder) and output (binary classification matrix)
- Container architecture (docker/kubernetes)
 - Free choice of development tools for participants
 - Safe for organizers
- Cloud computing: GCP/AWS or similar
- Challenge platform: crowdai.org/Kaggle etc.



kubernetes



Google Cloud Platform



kaggle

Performance metrics

	D disorders	D disorders																																										
Y^{true} : true test labels	<table border="1" style="border-collapse: collapse; width: 100%;"> <tr><td>1</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>1</td></tr> <tr><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>0</td></tr> <tr><td>1</td><td>0</td><td>1</td></tr> </table>	1	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	0	1	0	1	<table border="1" style="border-collapse: collapse; width: 100%;"> <tr><td>0</td><td>1</td><td>1</td></tr> <tr><td>0</td><td>0</td><td>1</td></tr> <tr><td>0</td><td>0</td><td>0</td></tr> <tr><td>1</td><td>1</td><td>1</td></tr> <tr><td>0</td><td>1</td><td>0</td></tr> <tr><td>1</td><td>1</td><td>0</td></tr> <tr><td>1</td><td>1</td><td>1</td></tr> </table>	0	1	1	0	0	1	0	0	0	1	1	1	0	1	0	1	1	0	1	1	1
1	1	1																																										
1	1	1																																										
0	0	0																																										
0	1	1																																										
1	1	1																																										
1	1	0																																										
1	0	1																																										
0	1	1																																										
0	0	1																																										
0	0	0																																										
1	1	1																																										
0	1	0																																										
1	1	0																																										
1	1	1																																										
	N subjects	N subjects																																										
	Y^{pred} : predicted labels																																											

Main metric (used for ranking): multi-task accuracy

$$\text{ACC} = 1 - \frac{1}{ND} \sum_{n=1}^N \sum_{d=1}^D |Y_{n,d}^{\text{true}} - Y_{n,d}^{\text{pred}}|$$

Secondary metrics: F1-score, sensitivity, specificity, precision, recall

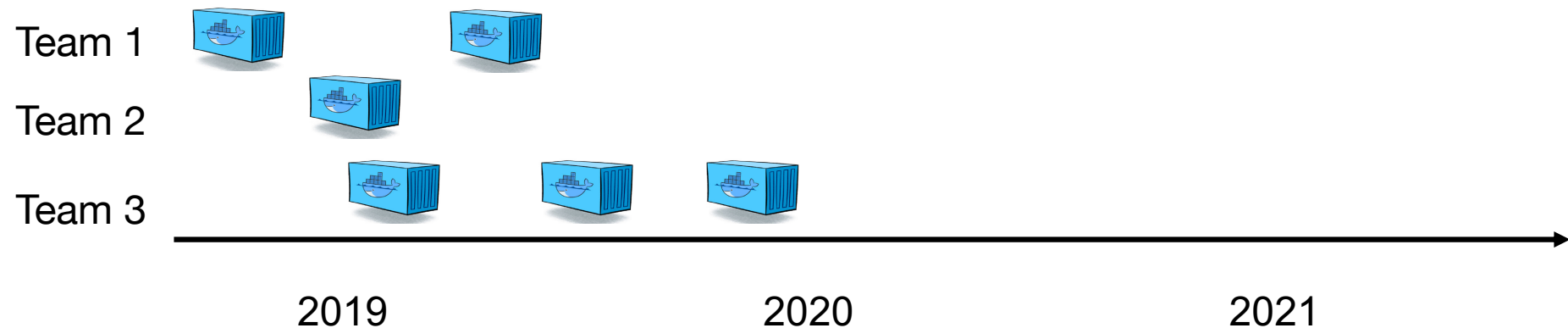
Multi-task metrics for continuous labels (severity scores) available.

Timeline

Idea: continuous prediction challenge

- Participant teams can refine and upload containers any time
- Benchmarking of most recent containers each time new data are released
- Time stamp system allows public release of test set without delay
- Tracking progress over time as new releases become available

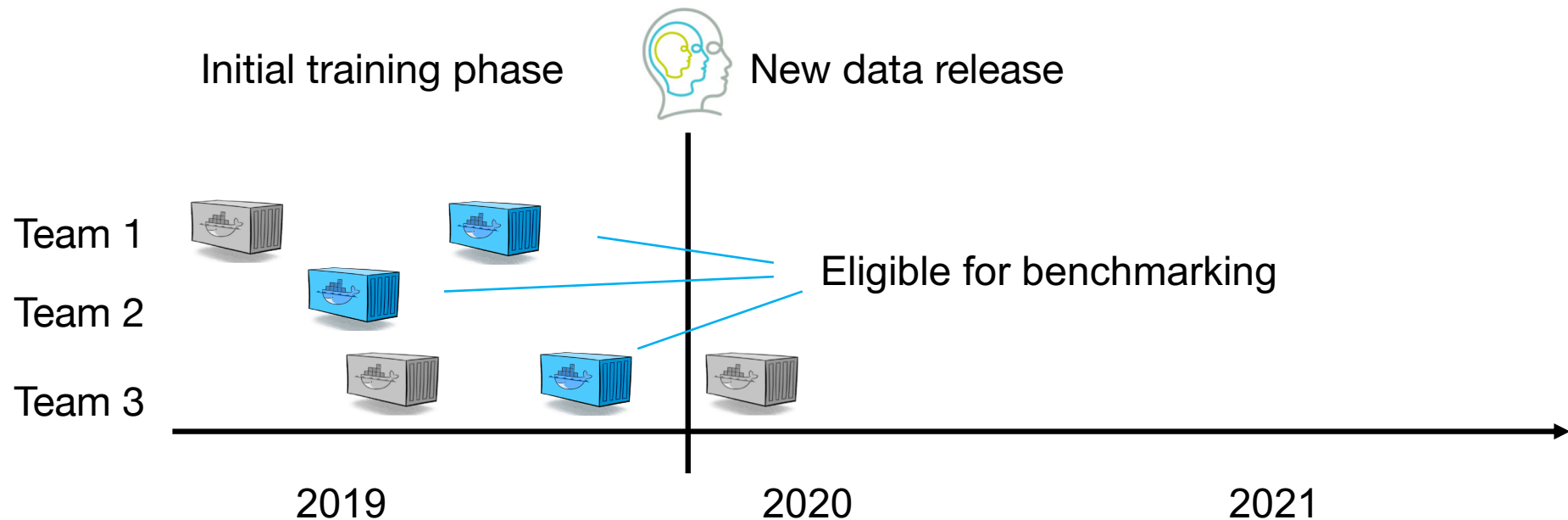
Initial training phase



Timeline

Idea: continuous prediction challenge

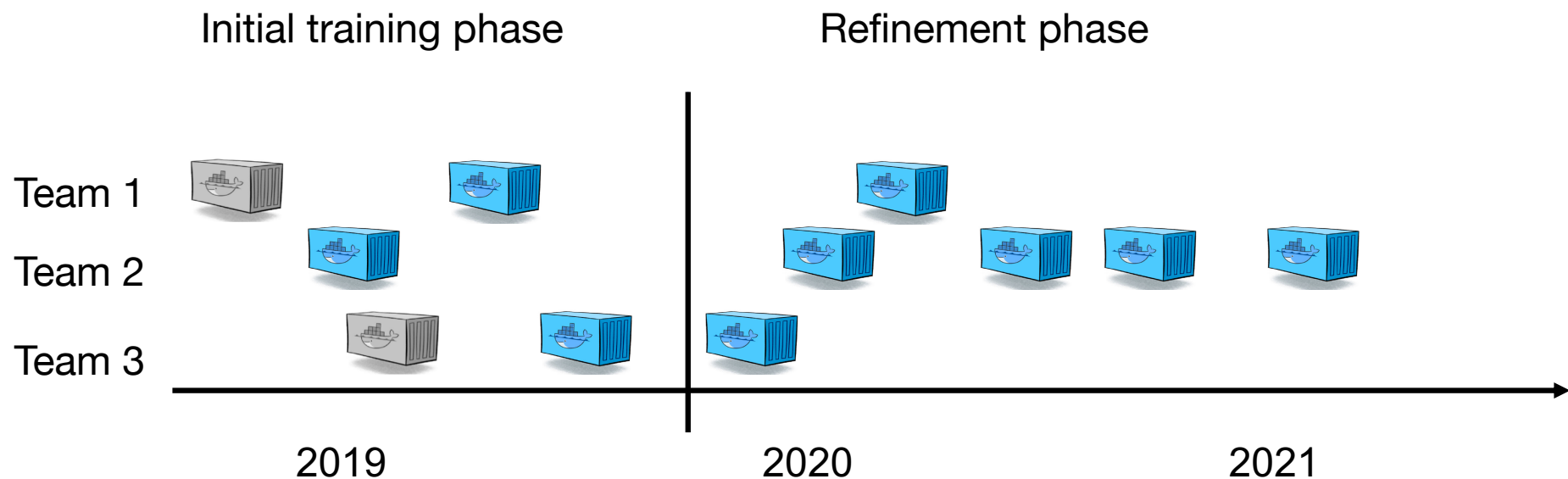
- Participant teams can refine and upload containers any time
- Benchmarking of most recent containers each time new data are released
- Time stamp system allows public release of test set without delay
- Tracking progress over time as new releases become available



Timeline

Idea: continuous prediction challenge

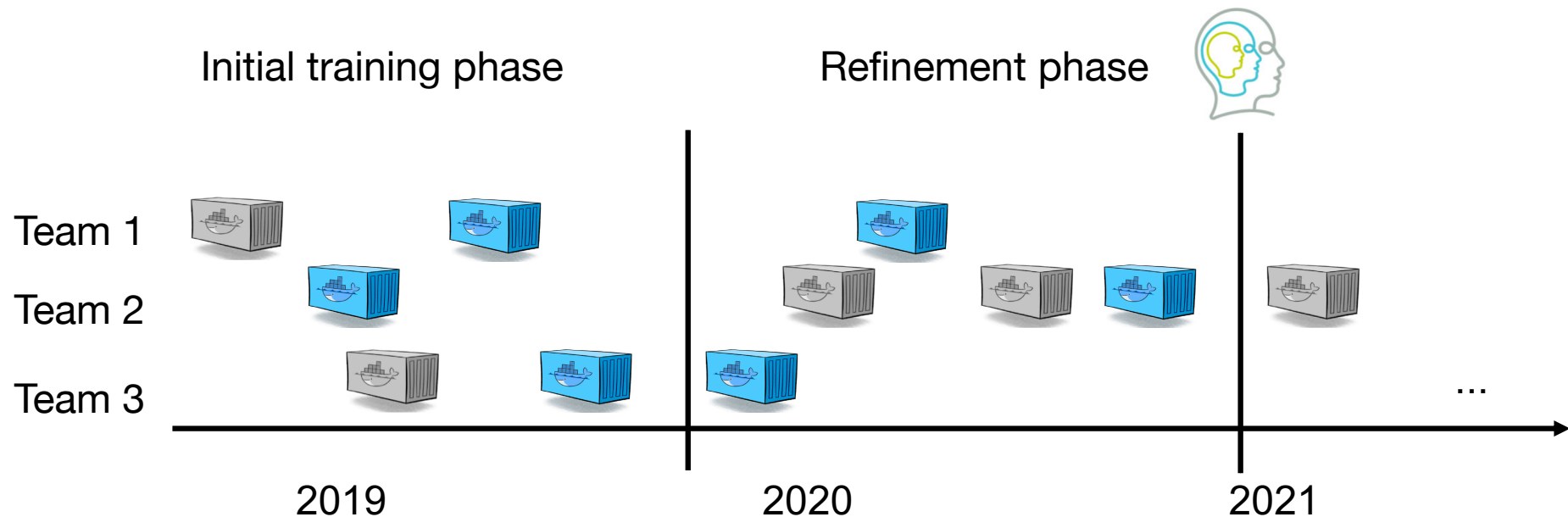
- Participant teams can refine and upload containers any time
- Benchmarking of most recent containers each time new data are released
- Time stamp system allows public release of test set without delay
- Tracking progress over time as new releases become available



Timeline

Idea: continuous prediction challenge

- Participant teams can refine and upload containers any time
- Benchmarking of most recent containers each time new data are released
- Time stamp system allows public release of test set without delay
- Tracking progress over time as new releases become available



Miscellaneous:

- Call for group participation (advertising on social media: Twitter)
- Work on C-105 document
- Infrastructure for data handling & management
- Feedback from psychiatry experts? (Noami Lee)
- Quantifying uncertainty

THANK YOU FOR YOUR ATTENTION