DELPHI SURVEY RESULTS

Utilizing a modified Delphi approach to select digital measures for common mental health disorders

DATACC
Digital Health Measurement Collaborative Community
by DiMe

COMMON MENTAL HEALTH DISORDERS
Digital Measures Development

View the resulting Conceptual Model

Set of Digital Measures of Common Mental Health Disorders

DATACC by DiMe

Project Team Approach

Meaningful Aspects of Health

Concepts of Interest (COIs)

Umbrella Review of Reviews

Review questions: What are the meaningful aspects of health to individuals with common mental health disorders? How is each concept meaningful to patients and under what contexts?

Outcomes to be Measured

Set of Digital Measures for Common Mental Health Disorders

Modified delphi approach *to identify appropriate outcomes for MH*

Selected a set of digital measures of CMHD, based on readiness for adoption, by reviewing:

- ***Maturity of technologies and measures*** *related to MH concepts of interest.*
- *Digital measures of MH being used in* ***clinical trials as endpoints*****.**
- ***Narrative review*** *of digital measurement for common mental health disorders, extending the previous* ***global research report****.*

Modified Delphi Approach to Identify a Consensus List of Digital Measures

A review of the literature and DiMe's Digital Endpoint Library was conducted to propose initial digital measures for common mental health disorders per concept of interest.

An electronic survey was sent to **subject matter experts (SMEs) in common mental health disorders and digital measurement to vote on measures for inclusion** and provide evidence and reasoning behind their selection.

Anonymized survey results were **summarized and shared with the SMEs**. The **survey was revised** accordingly and **sent back to the SMEs for re-vote.**

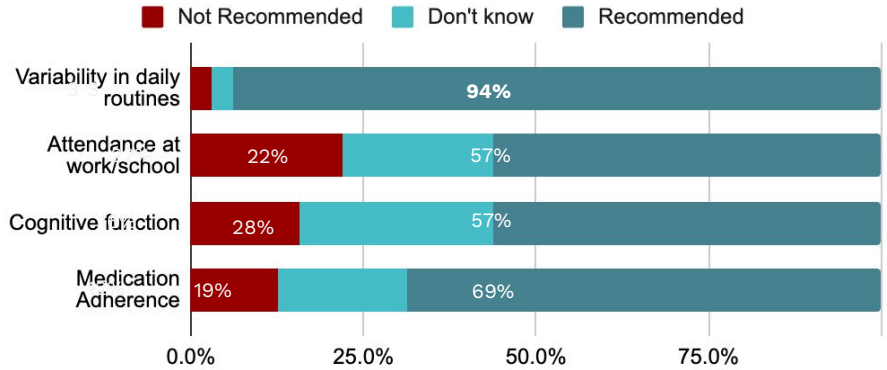
The **survey rounds** and anonymized results dissemination will continue **until consensus criteria was met**.

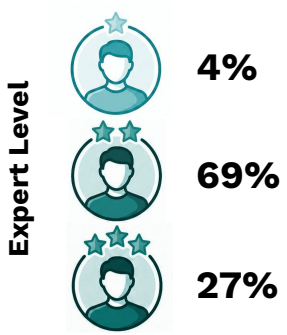
After no more rounds were necessary, there was a **live debrief workshop** to review measures selected and resolve remaining decisions related to inclusion/exclusion of measures.

Measures for **Routine Adherence**

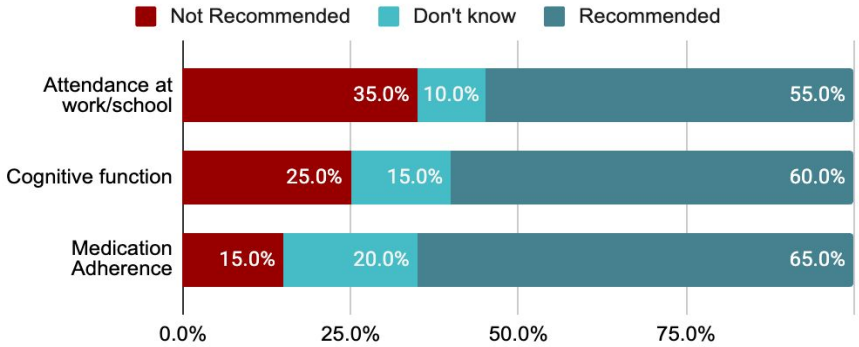
DATACC by DiME

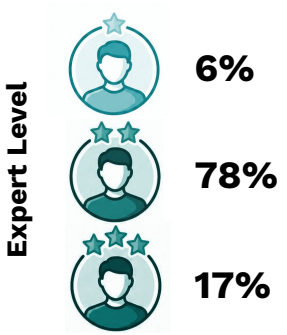
Round 1





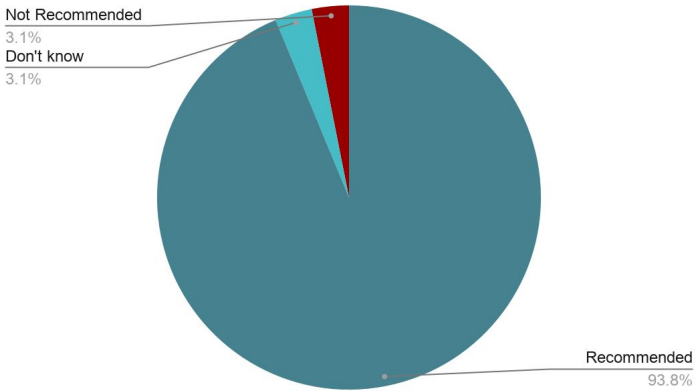
Round 2





Measures for **Variability in daily routines**

Round 1



Reported Reasoning for Inclusion:

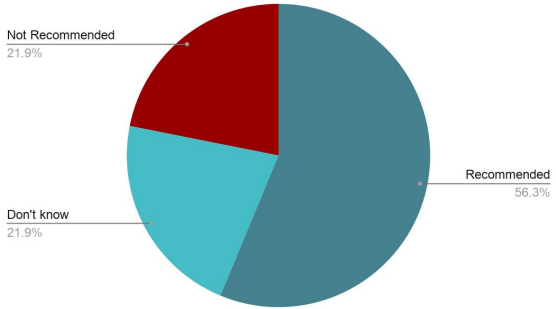
- Passive sensing measures like mobile apps capture real-world behaviors such as absenteeism and variability in daily routines. These are technically ready, clinically validated, and relevant across health indications, making them meaningful for assessing health aspects.
- I think variability within one's own daily routine reflects behavioral rhythm stability, is transdiagnostic and can captures change versus compliance.
- Mental health is largely driven by behavior and lifestyle choices. The more we can understand and quantify these outcome measures, the less 'fuzzy' disease endpoints will be
- simple routines can be taught even with diminished cognitive function, variability in daily routines seems like the most appropriate here.
- Changes in schedules are important indicators that something has shifted. It's highly objective but may require context.

Reported Considerations or Reasoning for Exclusion:

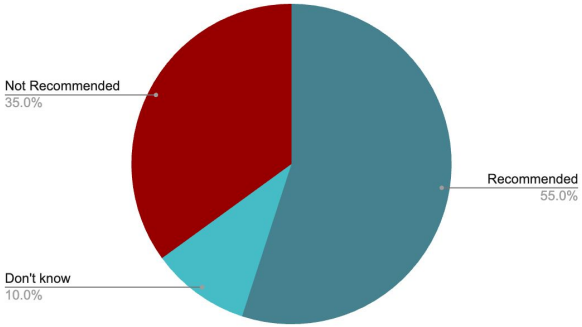
- these can be important factors/sources of variability to consider especially if there are ways to objectively measure them.
- For variability in daily routine, I would just want to note how different patterns might reflect positive change for different profiles...

Measures for **Attendance at work/school**

Round 1



Round 2



Reported Reasoning for Inclusion:

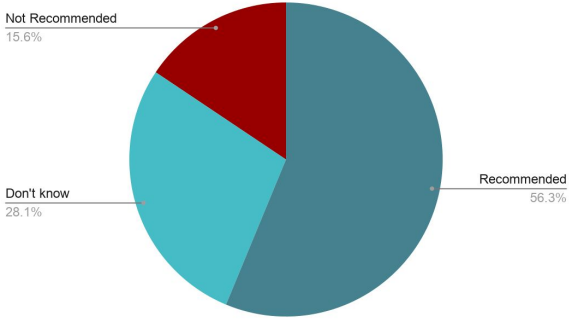
- All provide insight into the daily functioning which is a proxy to mental health functioning.
- while attendance at work/school has a correlation with mental health, there are a lot of considerations to be taken into account, including external factors that impact attendance (e.g., caregiving).
- Attendance in daily life activities is a global metric of impairment and functionality and seems to represent quotes from patients on what they view as living well.
- Time out of home is related to several mental health indices in our data, school attendance is a known correlate of depression and anxiety

Reported Considerations or Reasoning for Exclusion:

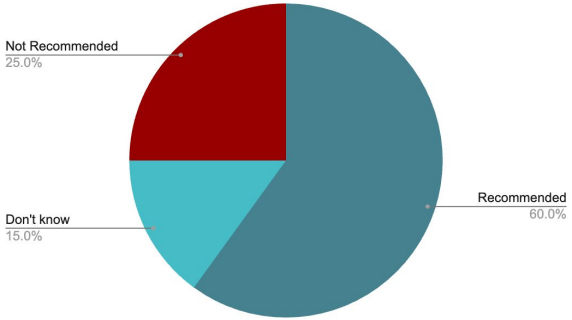
- Attendance of work/school might feel invasive as a measure, and it might not necessarily be related to mental health status and could be related to such a variety if external factors.
- All the selected measures are directly related to treatment effects **except for the school/work attendance** which is highly impacted by external factors outside of the pathology.
- I am not sure whether time in work/school is clearly relevant to mental health... so it's something I would drop out.
- Attendance needs to be contextual... broader exposure to social and enriching environments.
- From a privacy as well as a trust perspective, I would stay away from anything that touches work or school.
- I would not include... attendance at work/school as it is influenced by remote work, job type, caregiving duties, and possible SES constraints.
- While work and school could be conceivably supportive... in practice... these places do not support mental health well

Measures for **Cognitive function**

Round 1



Round 2



Reported Reasoning for Inclusion:

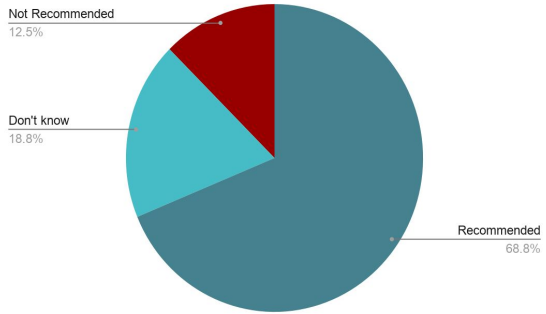
- Recommended due to the transdiagnostic indicators of attention, working memory and processing speed
- I think it would be hard to measure cognitive function in real world settings and because it can be influenced by so many other variables (stress, sleep, medications, etc
- Function and Cognition often go overlooked but are key aspects as mentioned and documented by patients
- Cognitive function... starting to show promise in characterizing and suggesting the presence of a mental health disorder
- Cognitive impairment is a core feature, not a side effect, of many mental health conditions.

Reported Considerations and/or Reasons for Exclusion:

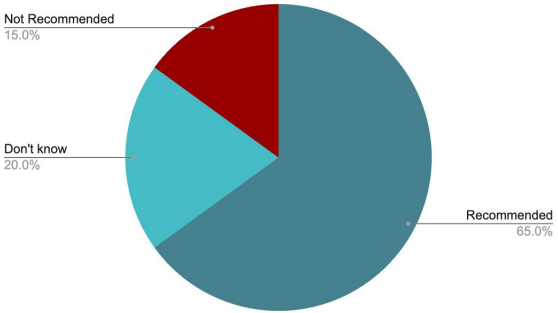
- For Cognitive function, I would be supportive of PVT and a match to sample task as they are don't take too much time...cogstate battery takes 15 minutes, which is far too long and should be excluded on that basis.
- Cognitive function is technically ready and validated... however, it is less suitable for mental health disorders...
- Cognitive function and its variability is hard to measure in real world settings.
- Cognitive testing would depend on how implemented. The more passively collected the better
- I would not include cognitive functioning only because it can be influenced by sleep, medications, stress, etc.
- The utility of these will always depend on the context of use... otherwise cognitive improvements can be hard to interpret
- Cognitive functioning is likely a secondary measure...
- I don't understand why cognitive function are in here

Measures for **Medication Adherence**

Round 1



Round 2



Reported Reasoning for Inclusion:

- ...important measure, especially in serious mental illness (bipolar, schizophrenia)
- Medication adherence technology is advanced...
- Poor adherence to medication regimen will help my treatment decision.
- Measuring medication adherence will help reducing the measurement noise.
- Medication adherence is a critical information
- Sleep & med adherence = paramount
- Medication adherence is the only measure listed that directly relates to adherence...

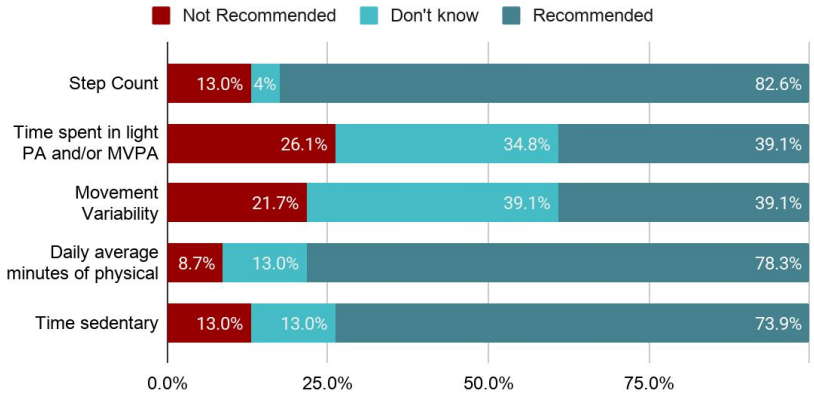
Reported Considerations or Reasoning for Exclusion:

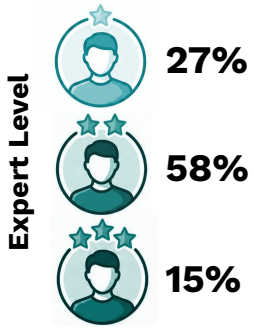
- ..not sure how accurately it can be measured and is not always relevant to all MH disorders
- Medications are not always prescribed for common mental health disorders, and in fact, often, behavioral therapies are first line of treatment depending on the diagnosis and severity.
- I do not believe medication adherence is an underlying mental health indicator because many people with mental health disorders do not take medication, and those who do may take it episodically or need routine use
- For medication adherence, is that referring to psychiatric meds or meds in general? there will be a lot of variability so I'm not sure how meaningful this will be unless its broken out more.
- Medication adherence is not a marker for underlying mental health
- Medication adherence could be used for a bipolar specifically perhaps
- It supports symptom improvement but *doesn't directly measure health outcomes.*
- I am not sure about medication adherence only because it can be treatment/disorder-specific...
- I tend to consider it a covariate or behavior of interest, not an endpoint.

Measures for **Movement**

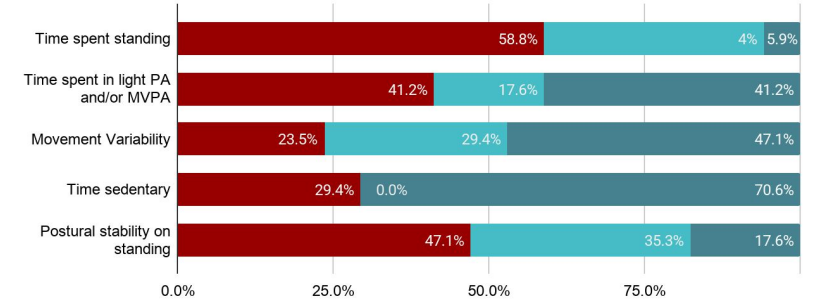
DATACC by DiME

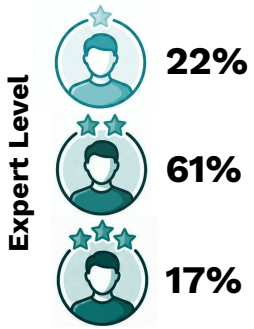
Round 1





Round 2





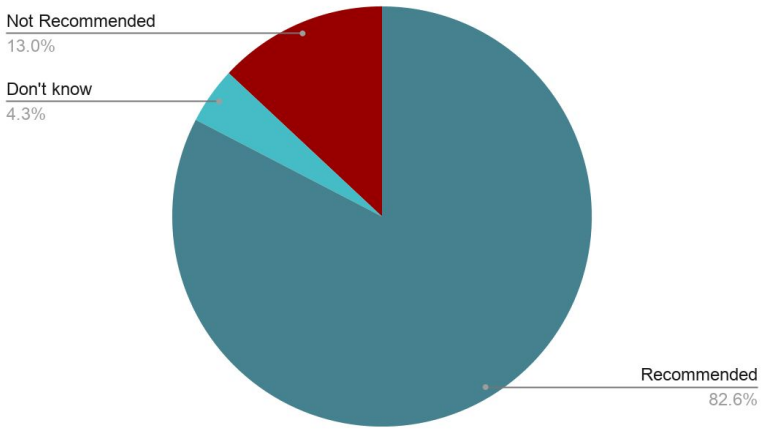
"We should recommend that they are baselined as opposed to generated with regards to norms that don't take into account lifestyle patterns etc."

"Daily movement can provide insight into variability of routine, level of enrichment, ability to engage in daily activities."

"The measures listed are not sensitive to health problems the patient may be experiencing."

Measures for **Step Count**

Round 1



Reported Reasoning for Inclusion:

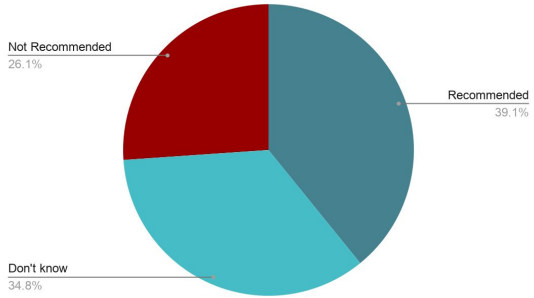
- If we had to choose between activities, number of steps and activity would provide information on the person overall mental health and have a certain amount of sensitivity
- Good Markers of global functionality, *exclude intense physical activity perhaps*
- Those data are accessible with basic wearables.
- Given what we know about exercise/movement and mental wellbeing, step count and/or physical activity could be meaningful correlates…

Reported Considerations or Reasoning for Exclusion:

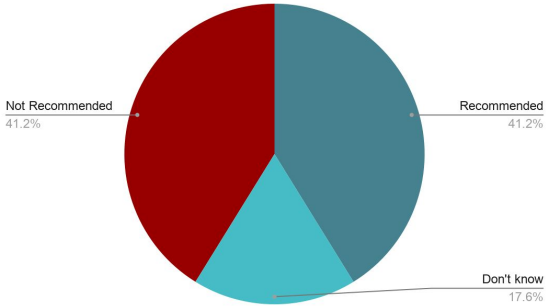
- Physical activity has a strong negative association with mental health symptoms. I don't believe step count adds much beyond time spent exercising…
- I am unsure/on the fence to include **step count**, time sedentary, and time spent in light PA or MVPA given the following factors that would need to be considered: people who are non-ambulatory, people whose job require sitting for long periods, and variation in health fitness among people.

Measures for **Time spent in light PA / MVPA**

Round 1



Round 2



Reported Reasoning for Inclusion:

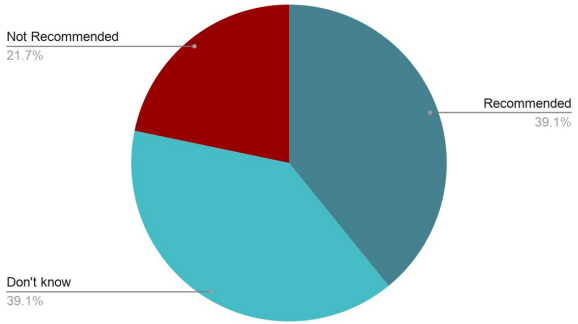
- MVPA pops up across many therapeutic areas and has a precedent for use as a primary digital endpoint for iPF
- Time sedentary and activity seem like the most important and feasible outcomes of the group
- We collect these measures longitudinally at [Company] and they have guided prescription decisions

Reported Considerations or Reasoning for Exclusion:

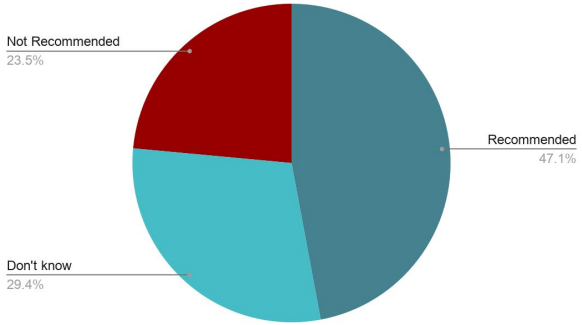
- Those data are accessible with basic wearables. Time spent in light [PA] unavailable on iOS i think
- I know that MVPA has been accepted as a measure for iPF and it pops up as important for other TAs e.g. core measures of physical activity.
- Time spent in light to MVPA I voted no, because it can be also a coping mechanism for people when their condition worsens.
- I would caveat on for both time spent in light to MVPA that it is *highly dependent on an individual's situation* and existing patterns... For individuals going through a busy work period, or who have generally sedentary jobs, they may end up showing lower movement through the day, but that might not be representative of their actual mental health state.
- I am unsure/on the fence to include step count, time sedentary, and **time spent in light PA or MVPA** given the following factors that would need to be considered: people who are non-ambulatory, people who's job require sitting for long periods, and variation in health fitness among people.

Measures for **Movement Variability**

Round 1



Round 2



Reported Reasoning for Inclusion:

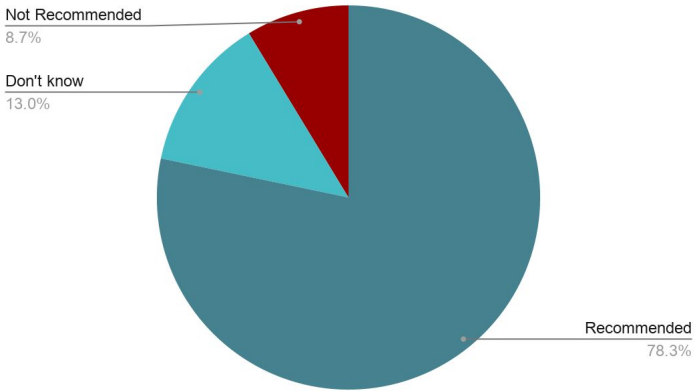
- Movement variability and postural stability seem like measurements that would be more suitable to only a subset of MH conditions, and not all
- Given the strong correlation between movement and wellness, these *all* seem very reasonable.

Reported Considerations or Reasoning for Exclusion:

- Based on literature evidence> I would not recommend Movement Variability as it's not easy to capture nor to define quantitatively
- Movement Variability seems like something that can be caused by various comorbidities rather than just mental health.
- Movement variability tends to be more associated with neurological disorders.

Measures for **Daily average minutes of physical activity per time**

Round 1



Reported Reasoning for Inclusion:

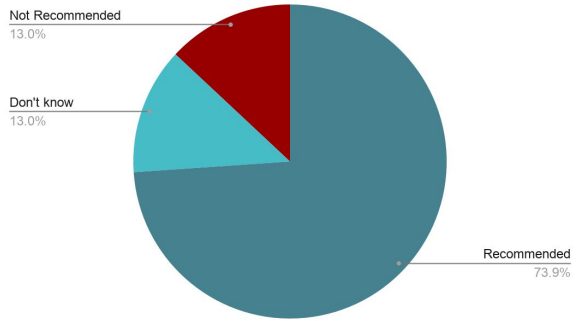
- Daily average minutes of physical activity is an objective, scalable indicator of real-world functioning in mental health...
- We collect these measures longitudinally at [Company] and they have guided prescription decisions
- These measures capture intermediate aspects of movement as well which is useful
- Given what we know about exercise/movement and mental wellbeing, step count and/or physical activity could be meaningful correlates of wellbeing. I would want to take health concerns and mobility concerns into account, so perhaps considering ability status and using a *broad definition of physical activity* would be most broadly useful.
- Physical Activity is a core health measure

Reported Considerations or Reasoning for Exclusion:

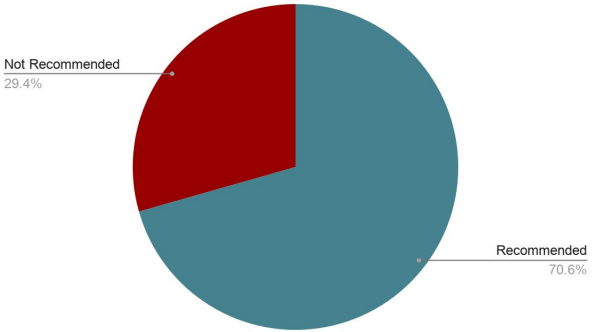
- Anhedonia and apathy are key to several psych diseases [recommended Daily average minutes PA/time & Time Sedentary]
- There is a lot of research to suggest that physical activity has a measurable impact on mental health.

Measures for **Time sedentary**

Round 1



Round 2



Reported Reasoning for Inclusion:

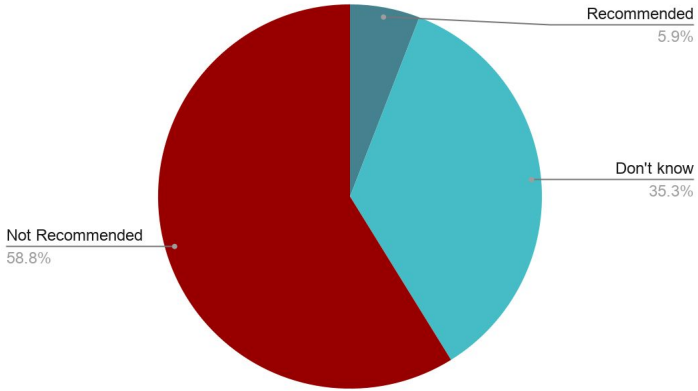
- Time spent standing and sedentary (during awake time I suppose) would add to 100% of the awake time, so are basically two coins of the same measure.
- The only measure I would lean toward including is sedentary time due its correlates with mental and physical health.
- Upon further reflection, I do think time sedentary can compliment "step count."
- ...only the sedentary time would be relatively reliable from self report.
- Sedentary time also important, but not yet precedent for primary endpoint.
- Measure changes in sedentary time alone might be the most valuable single metric when it comes to activity.
- Movement or lack thereof is highly related to assessing mental health stability.
- **Time sedentary** and activity seem like the most important and feasible outcomes of the group

Reported Considerations or Reasoning for Exclusion:

- Time spent standing and sedentary time seems very dependent on job/caregiving tasks and physical ability and with those limitations..
- I'm concerned that sedentary time and light exposure are not sufficiently discriminative.

Measures for **Time spent standing**

Round 2



Reported Reasoning for Inclusion:

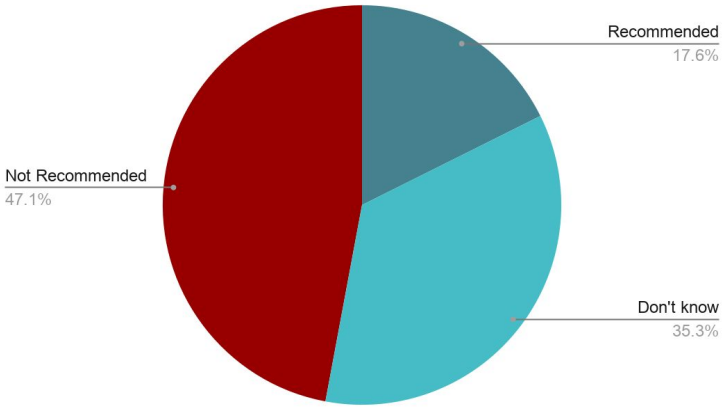
- Time spent standing and sedentary (during awake time I suppose) would add to 100% of the awake time, so are basically two coins of the same measure.

Reported Considerations or Reasoning for Exclusion:

- Time spent standing can have a lot of unexplained variability...
- In our analysis we don't find time standing to be strongly associated with depression and anxiety, we do find associations with physical activity.

Measures for **Postural stability on standing**

Round 2



Reported Reasoning for Inclusion:

- Movement variability and postural stability seem like measurements that would be more suitable to only a subset of MH conditions, and not all
- If there is a way to capture posture at home objectively, I would suggest that be included.

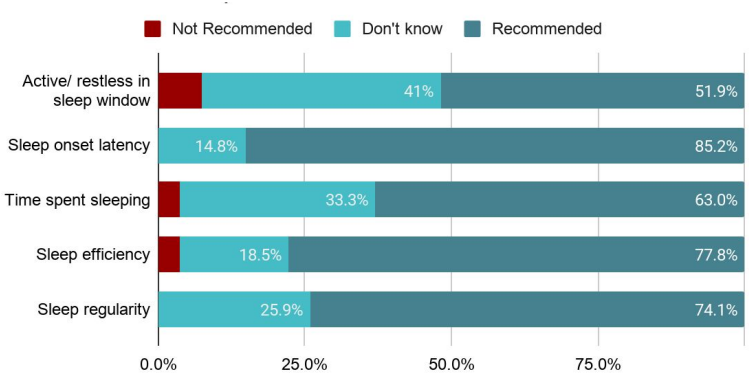
Reported Considerations or Reasoning for Exclusion:

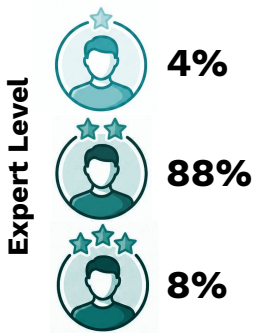
- Upright/posture stability - I don't know how this relates to common mental health disorders - so I vote to exclude.
- While movement and activity are good indicators of mental state, not sure about the standing questions.

Measures for **Sleep**

DATACC by DiME

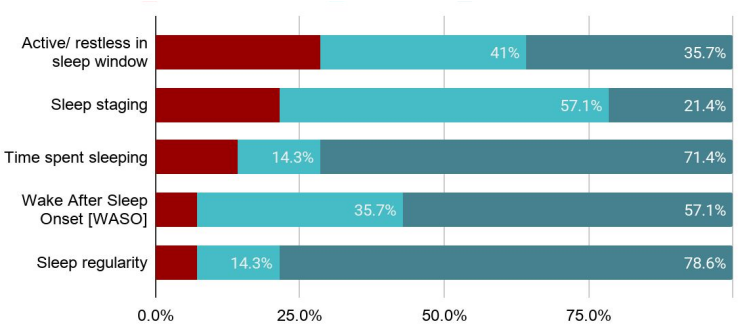
Round 1

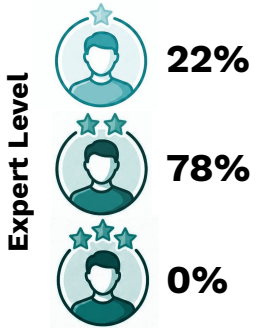




"All these measures are related to mental health outcomes and affected by mental health and mental health treatment adherence."

Round 2

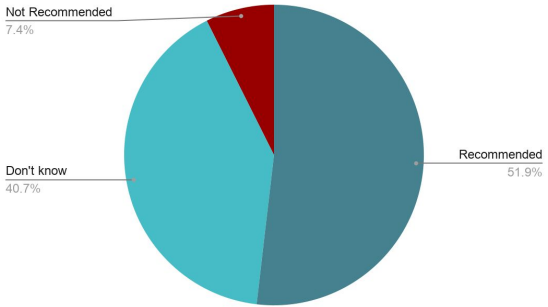




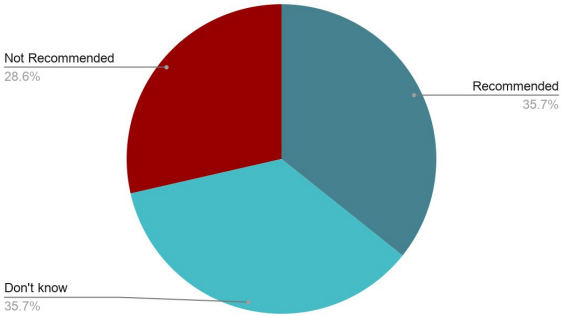
"Sleep can be a strong indicator of mental health and also relatively easy to measure objectively."

Measures for **Active / restless in sleep window**

Round 1



Round 2



Reported Reasoning for Inclusion:

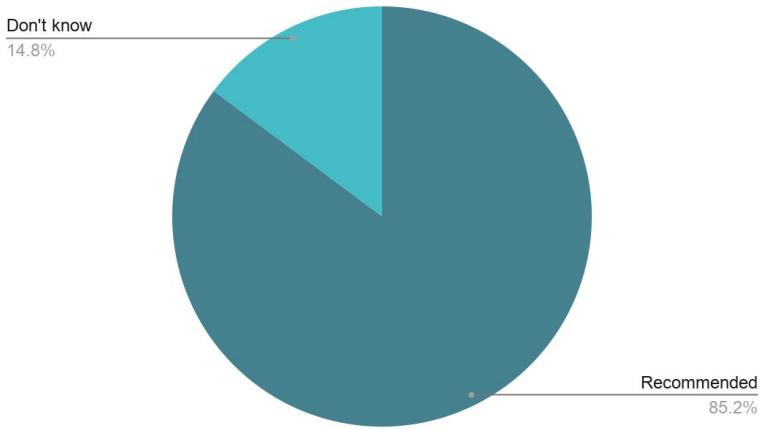
- Regularity more likely to be influenced by non MH factors; same with active/restless sleep
- The first four measures are items which have been shown to be altered in individuals with mental health disorders, *but also are items which can show differences day to day regardless of time spent sleeping.*

Reported Considerations and/or Reasoning for Exclusion:

- ...sleep staging and restlessness could be affected by many factors and also administered therapies more likely than those I selected, therefore I suggest their exclusion
- I do think there's quite a bit that can be gleaned from sleeping patterns, but, as mentioned by others in the previous round, active/restless periods in sleep window can be caused by a lot of other issues other than the mental health disorder.
- There would be a lot of value in adding sleep staging, purely behavioral measures can be limiting.
- I do not think **activity during sleep** and sleep efficiency are good measures because of the following confounding factors: (1) sleep disorders and/or neurological conditions, (2) people spend time in bed awake doing other activities, and (3) *unclear how sleep efficiency relates to disorders outside of anxiety/insomnia*.
- Not selected Sleep onset latency (time) and **Active/ restless in sleep window** purely because can have other causes rather than changes in mental health state.

Measures for **Sleep onset latency**

Round 1



Reported Reasoning for Inclusion:

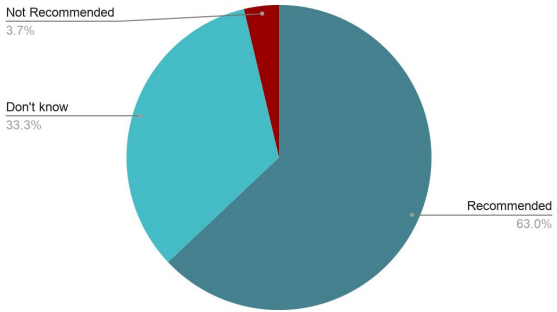
- Sleep onset and Sleep efficiency are widely used in RCTs, WASO should be added also.
- Sleep latency has been useful in our longitudinal work *(Not passive sensing)*
- ..[sleep] regularity is important but could also be captured by onset latency probably

Reported Considerations or Reasoning for Exclusion:

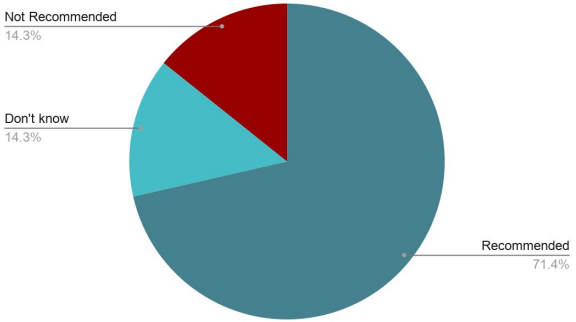
- Not selected **Sleep onset latency** ... purely because can have other causes rather than changes in mental health state. Sleep regularity can be influenced by many environmental & external factors.
- I am unsure about sleep onset latency given it is closely related to cognitive arousal/rumination and may be more specific to stress, anxiety and insomnia; it's also dependent on when bedtime is inferred.

Measures for **Time spent sleeping**

Round 1



Round 2



Reported Reasoning for Inclusion:

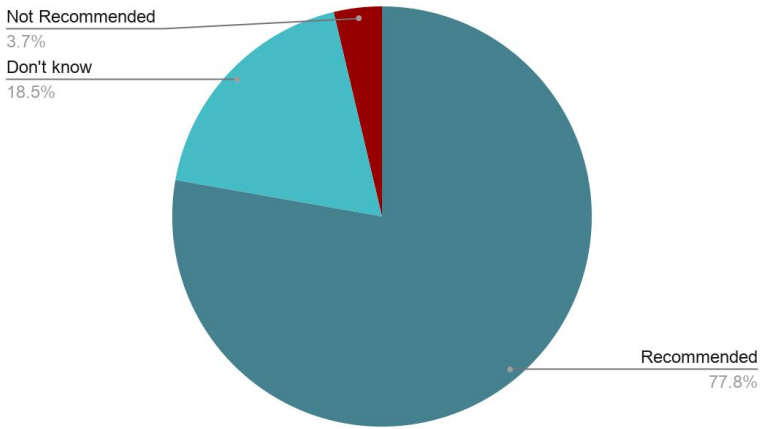
- My experience with passive sensing and youth mental health finds links between overall sleep time (with too long or too short related to adolescent depression/anxiety) as well as time to fall asleep as important
- Disruptions in sleep routine and duration can be an early warning sign for a change in the intensity/significance of mental health disorders.
- I like time spent sleeping overall because it is transdianogistc and it should focus on within-person changes *(though note, sleep duration does not equal sleep quality*).
- I think time is a must have, as variations in # of hours slept can be huge in mood disorders.
- The first four measures... add additional color to the raw number that would be obtained in time spent sleeping
- Sleep metrics such as onset latency, efficiency, duration, fragmentation, and regularity capture a foundational biological process that is transdiagnostic, predictive of relapse and functional outcomes, and mechanistically linked to cognition, mood regulation, and daily activity.
- Total time spent sleeping (not too much or not too little) is important for mental health.

Reported Considerations or Reasoning for Exclusion:

- I am on the fence about total time sleeping because while transdiagnostic and easy to measure, sleep duration does not always equate to sleep quality.
- Sleep regularity seems to be more associated with well being than just hours slepts
- total time may not be useful, as MDD can cause people to oversleep, or undersleep.

Measures for **Sleep efficiency**

Round 1



Reported Reasoning for Inclusion:

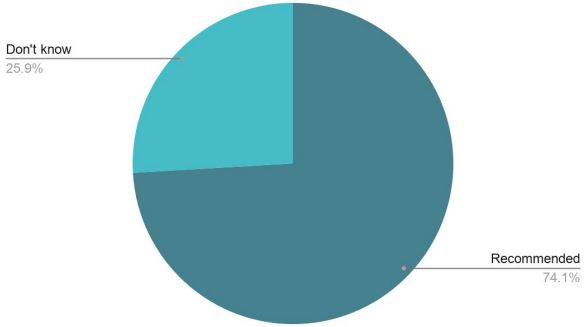
- Sleep efficiency tends to be the measure that is **most meaningful** across mental health disorders, compared to the other listed.
- Sleep quality can help determine what better explains mental health symptoms.
- Efficiency can be more important than actual time of sleep
- Sleep onset and Sleep efficiency are widely used in RCTs
- Sleep metrics such as onset latency, **efficiency**, duration, fragmentation, and regularity capture a foundational biological process that is transdiagnostic, predictive of relapse and functional outcomes, and mechanistically linked to cognition, mood regulation, and daily activity.

Reported Considerations or Reasoning for Exclusion:

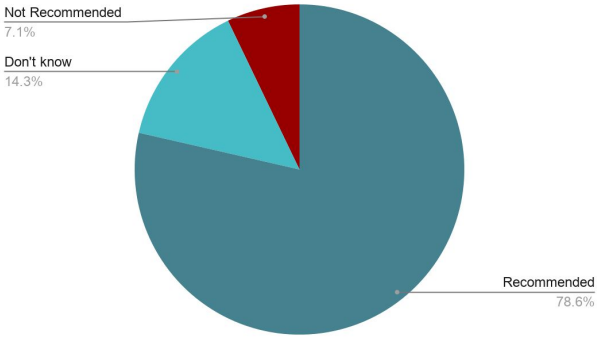
- Individual changes in these parameters will be most important as opposed to their values. [Recommended]
- I do not think activity during sleep and **sleep efficiency** are good measures because of the following confounding factors: (1) sleep disorders and/or neurological conditions, (2) people spend time in bed awake doing other activities, and (3) unclear how sleep efficiency relates to disorders outside of anxiety/insomnia.

Measures for **Sleep regularity**

Round 1



Round 2



Reported Reasoning for Inclusion:

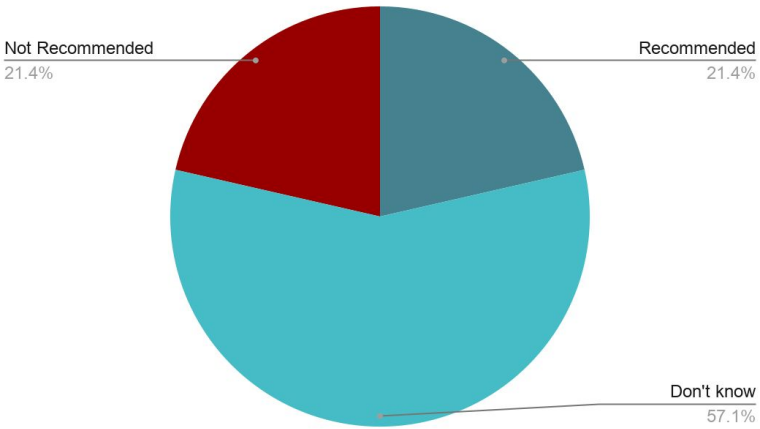
- While these measures can change due to factors outside of mental health concerns (e.g., health issue, pregnancy), those are also periods of increased vulnerability to mental illness onset so I don't think that is inherently problematic.
- Sleep regularity and WASO can be helpful to measure sleep routine/fragmentation.
- Sleep regularity has been found to be most sensitive measure of sleep to detect if someone is unwell.
- Sleep regularity seems to be more associated with well being than just hours slepts
- Sleep regularity is important for overall wellbeing, especially in the context of bipolar disorder.
- Sleep regularity can be predictive of the onset of manic symptoms.
- I also think sleep regularity is transdiagnostic and closely linked to circadian stability and mood regulation; however, the caveat here is that you do need to take into consideration shift workers and caregivers...
- if people's sleep patterns are very erratic, that is a sign of instability in their life/routine (as you might have in bipolar, or in Emotionally Unstable / Borderline Personality Disorder

Reported Considerations or Reasoning for Exclusion:

- Regularity more likely to be influenced by non MH factors; same with active/restless sleep
- Regularity is important but could also be captured by onset latency probably
- Sleep regularity can be influenced by many environmental & external factors.

Measures for **Sleep staging**

Round 2



Reported Reasoning for Inclusion:

- Sleep regularity seems to be more associated with well being than just hours slepts

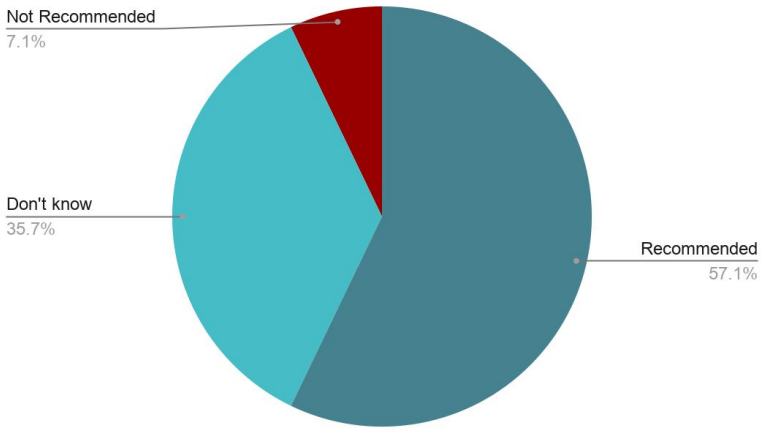
Reported Considerations or Reasoning for Exclusion:

- ..sleep staging and restlessness could be affected by many factors and also administered therapies more likely than those I selected, therefore I suggest their exclusion
- I don't think sleep staging is yet accurate enough by wearables to add reliable objective information

Measures for **Wake After Sleep Onset [WASO]**

DATACC by DiME

Round 2



Reported Reasoning for Inclusion:

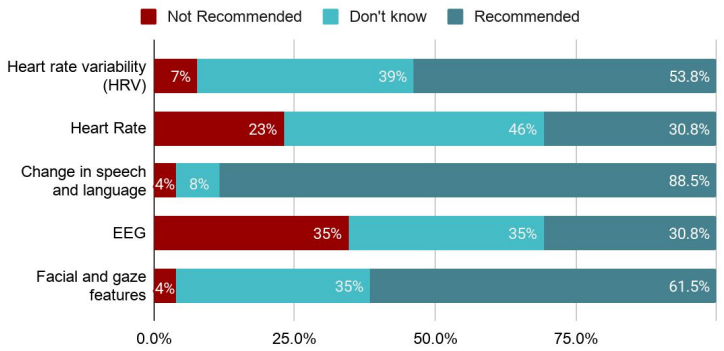
- Sleep regularity and WASO can be helpful to measure sleep routine/fragmentation.
- WASO has popped up alot in literature.

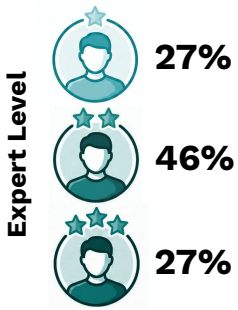
Reported Considerations or Reasoning for Exclusion:

- Regularity is important but could also be captured by onset latency probably

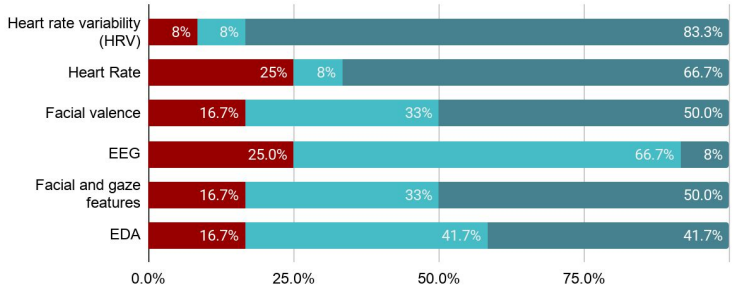
Measures for **Frequency, intensity, & change in emotions**

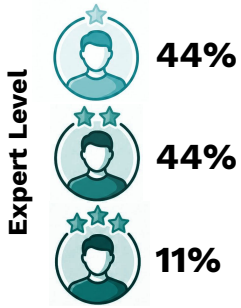
Round 1





Round 2



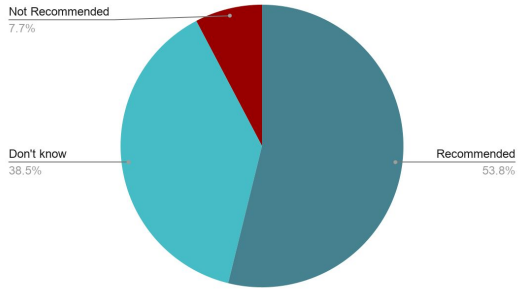


"There's lot of evidence to support physiological measures as being important for assessing mental health."

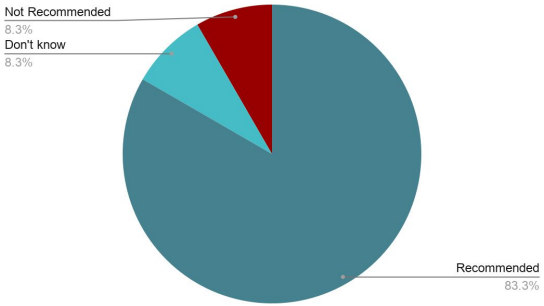
"They are technically complex, poorly standardized across settings, and often have unclear clinical meaning compared with simpler, widely validated symptom and functioning scales."

Measures for **Heart rate variability (HRV)**

Round 1



Round 2



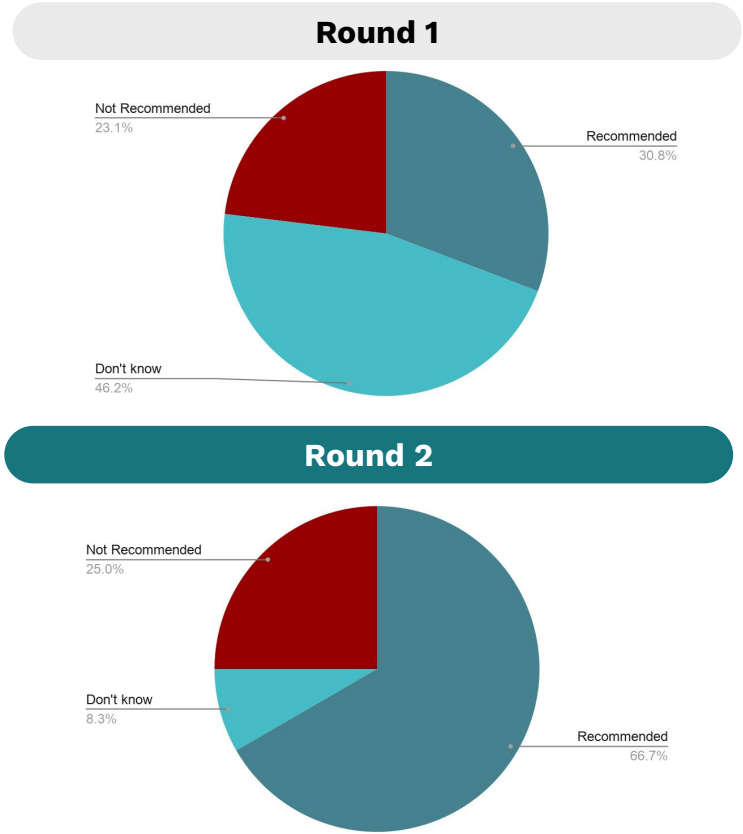
Reported Reasoning for Inclusion:

- HRV, EDA and EEG all are developing measures for common MHD
- HRV links to a number of coping, stress, and emotion measures in ESM in our data
- We've seen a lot with HRV and mental health generally
- With HRV, it appears to provide more information than just raw heart rate counts.
- Heart rate and **HRV** are strongly linked to mental health disorders and technology to measure these variables exist and can be done at home.
- HRV is quite transdiagnostic with a body of evidence supporting its clinical validation across multiple mental health disorders and it is tech ready.
- *I know that there is plenty of **emerging data** regarding speech characteristics and emotional states. Similar with **HRV** and facial features.*
- HR/HRV can add another dimension (passive measurement).

Reported Considerations and/or Reasoning for Exclusion:

- I think many of these measures (HRV, EEG, etc) are not feasibly enough to be measured at this point in time and lack sufficient evidence of effectiveness
- I am on the fence on HRV given it's strong correlation to disorders and yet, can be influenced by things like caffeine.
- HRV is *emerging* as important in a number of TAs as an important measure of neurological status.
- I haven't seen any convincing technologies for reliably measuring emotion in any of these ways. **HRV** isn't reliable – the actual technology used... can lead to variability...
- I don't think we currently know enough about any measure of HRV and what it adds to just HR data.
- While they are all important and highly discriminative, I think they must be done in an unobtrusive, ubiquitously available way. That rules out EEG and makes HR/**HRV** quite hard to include without relying on less-ubiquitously available hardware such as wearables.

Measures for **Heart rate**



Reported Reasoning for Inclusion:

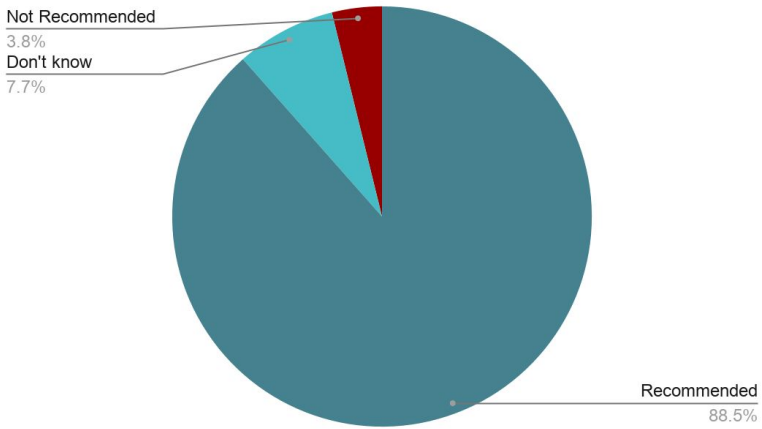
- In our analyses we find strong associations between heart rate and facial emotion recognition and depression and anxiety symptoms
- HR I think could be very helpful when looked at in response to specific tagged stimuli- but you'd need the context for it.
- Heart rate is extremely important, and numerous papers show it can be predictive of mental health symptoms
- **Heart rate** and HRV are strongly linked to mental health disorders and technology to measure these variables exist and can be done at home.
- **HR**/HRV can add another dimension (passive measurement)
- I feel very confident about heart rate (*assuming heart rate is personalized*...) given its relevance for trauma and anxiety.

Reported Considerations or Reasoning for Exclusion:

- Cardiovascular measures need to be considered along with other height and weight measures
- Heart beat & EEG : less evidence on direct correlations with symptoms?
- Similar to previous sections, with Heart Rate, based on research, the measures that will be the most informative will be additional information that can be gained by looking further into the HR wave, which is captured by HRV and hopefully other measures that can be *gathered that go beyond raw* **heart rate.**

Measures for **Change in speech & language**

Round 1



Reported Reasoning for Inclusion:

- Speech: it's what mental health clinicians analyse already in their today's practice
- Language style seems to have potential in relation to personality disorders and anxiety disorders in teens
- Speech rate would be useful for both depression and mania assessment.
- Speech features are exploratory but a growing body of literature demonstrates the relevance of speech features to monitor symptoms (e.g., pitch variability in MDD) and potentially support patient selection strategies (e.g., speech latency in bipolar depression).
- The positive or negative valence of words would also be meaningful.
- I know that there is plenty of emerging data regarding speech characteristics and emotional states.
- Speech is a rising marker and would fit well.
- There are a lot of data showing that voice features are associated with well being. However, it may be less feasible to measure

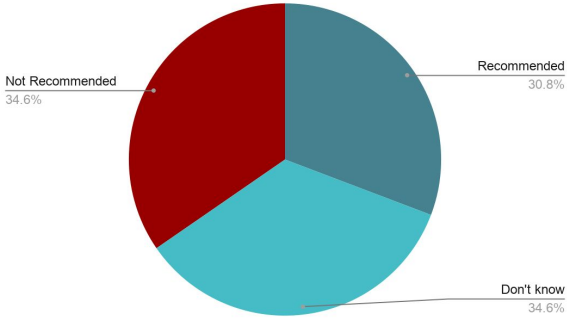
Reported Considerations or Reasoning for Exclusion:

- Regarding speech metrics, I think they could be valuable, but the state of the art is not yet advanced enough...
- While I like the change in speech/language... it is also highly language-, culture-, and context-dependent...

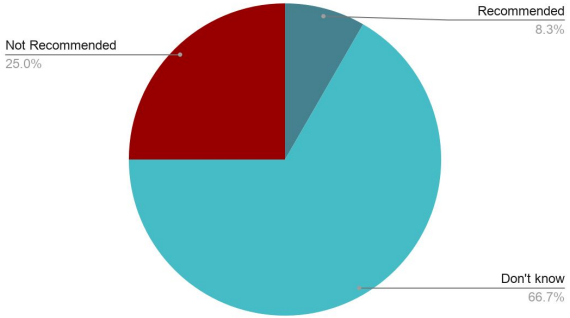
Features from **Electroencephalogram (EEG)**

DATACC by DiME

Round 1



Round 2



Reported Reasoning for Inclusion:

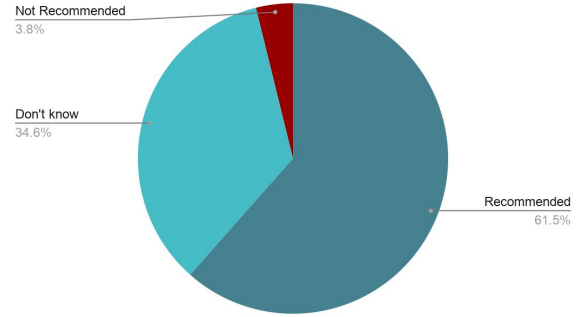
- HRV, EDA and EEG all are developing measures for common MHD
- There is a substantial and growing evidence base supporting the inclusion of EEG...
- EEG is particularly important... because it provides a direct measure of neural activity...
- Yes familiar with evidence for each feature in research. ERP for EEG

Reported Considerations or Reasoning for Exclusion:

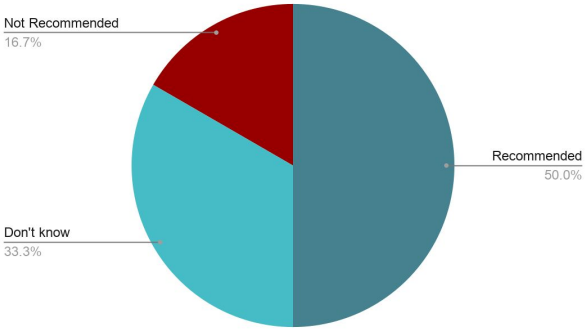
- EEG I said no due to feasibility in the wild, but again, if context was specified I feel like lab experiments do demonstrate strong links to error dectection/anxiety.
- I think many of these measures (HRV, EEG, etc) are not feasibly enough to be measured at this point in time and lack sufficient evidence of effectiveness
- EEG seems unreasonable as a measure -- collecting/interpreting.
- EEG, I still think it's too invasive.
- EEG is more difficult to scale.
- Difficult to get everyone an EEG...
- With EEG, capture techniques and movement... could affect the outputs...
- I don't think we currently know enough... EEG are nice but too cumbersome.
- EEG hard to operationalize also don't know how much can reflect emotional states.
- EEG does not seem like a viable option considering adoption and limited use outside of labs.

Measures for **Facial & gaze features**

Round 1



Round 2



Reported Reasoning for Inclusion:

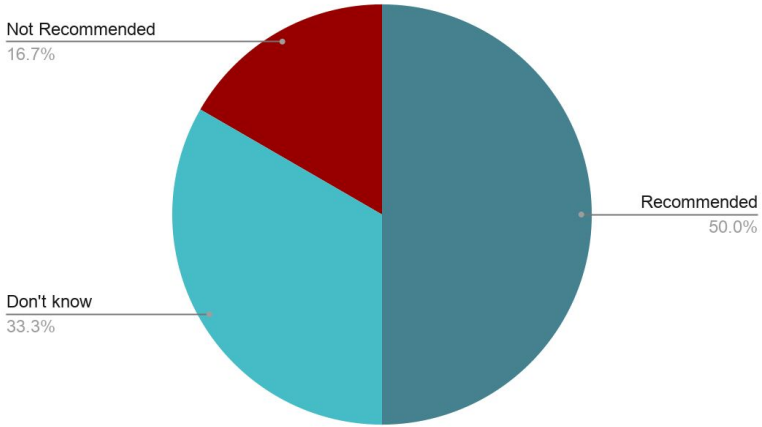
- Facial features, while potentially leading to privacy concerns, have shown promise in individual baselining for mental health disorders.
- Facial features are also associated with these symptoms and provide a good assessment of a person's emotions.
- Not recommending, but facial recognition so far has been the best (digital) measure of emotions.
- Difficult to get everyone an EEG but the facial features and changes in emotion are key for anhedonia and psychomotor retardation
- I know that there is plenty of emerging data regarding speech characteristics and emotional states. Similar with HRV and facial features.
- Facial and gaze features would be the best measure, but harder to operationalize.
- I'd like to suggest that facial valence be used as this has been validated in many studies

Reported Considerations or Reasoning for Exclusion:

- Consider the invasiveness of videos and what it would be like to utilize facial/gaze tracking in practice.
- Eye tracking I'm only familiar with the autism literature, and that might not be 'transdiagnostic enough'.
- Not sure about the level of evidence for Facial valence and Gaze tracking.
- Facial features and gaze patterns are still quite experimental...
- These measures tend to require capture of patient faces that can raise privacy concerns.
- Video way too invasive.
- Facial gaze, etc... are nice but too cumbersome
- I'm a little unsure of the validation status of facial and gaze features...

Measures for **Facial valence**

Round 2



Reported Reasoning for Inclusion:

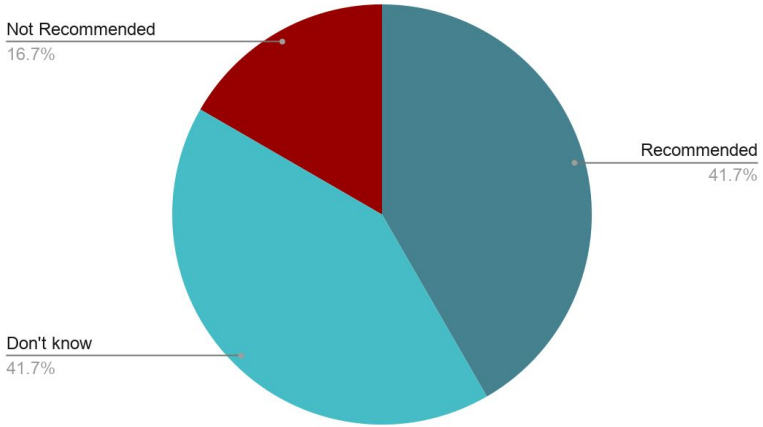
- As mood disorders, capturing objective information about valence and arousal is highly valuable.

Reported Considerations or Reasoning for Exclusion:

- Facial valence I feel like is typically a wash in the literature.

Measures for **Electrodermal activity (EDA)**

Round 2



Reported Reasoning for Inclusion:

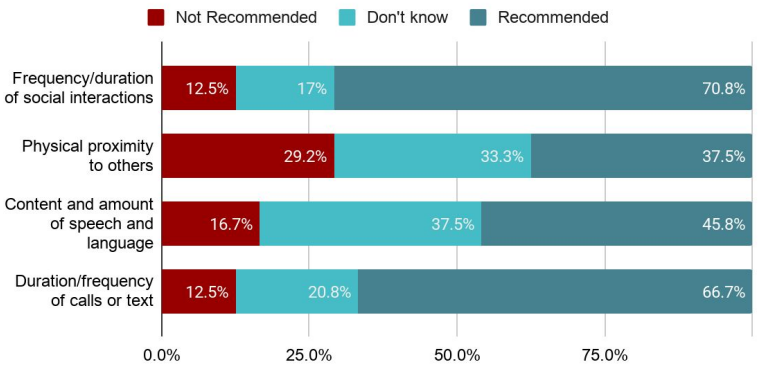
- HRV, EDA and EEG all are developing measures for common MHD

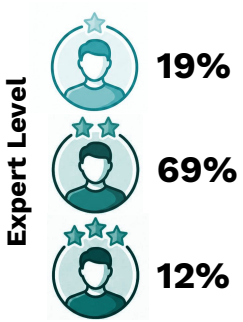
Reported Considerations or Reasoning for Exclusion:

- While EDA and HRV theoretically track ANS activity, I don't think they are adequately validated as meaningful.

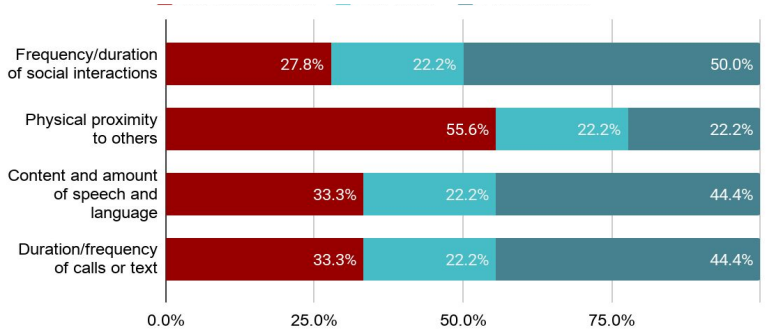
Measures for **Change in social interactions**

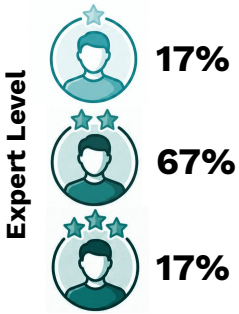
Round 1





Round 2





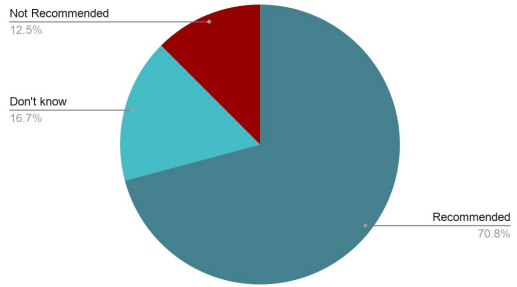
"Some of these measure may be do well when in a larger model but on their own are not sufficiently sensitive and specific."

"Lack of specificity. location-based measures are very context based"

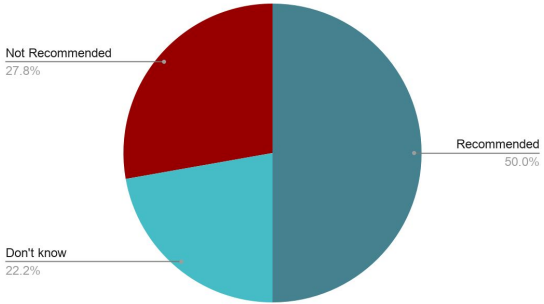
"These can be a proxy for social engagement, which can reflect changes in severity of mental health disorders."

Measures for **Frequency/duration of social interactions**

Round 1



Round 2



Reported Reasoning for Inclusion:

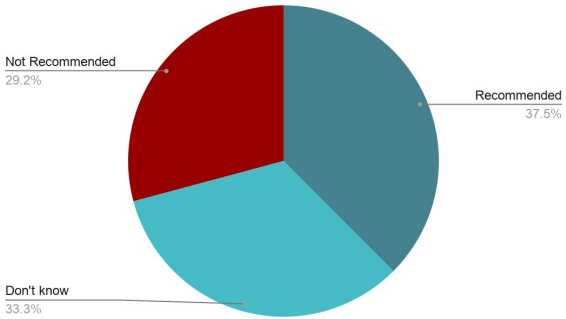
- I am not sure how specific the measures of social interaction are, but they seem reasonably sensitive to mental health that it's worth including/considering.
- Another measure of activity. The more lifestyle measures we can cover without burdening the patient will give us a clearer view
- HRV is quite transdiagnostic with a body of evidence supporting its clinical validation across multiple mental health disorders and it is tech ready.
- Social interaction and social networks probably play a large role as well. again, *not sure if they're core or if they can provide additional information*
- Frequency/duration of social interactions fits best if context ix captured
- I'd recommend these but with analyze data with caution. Controlling for other variables will be crucial here

Reported Considerations and/or Reasoning for Exclusion:

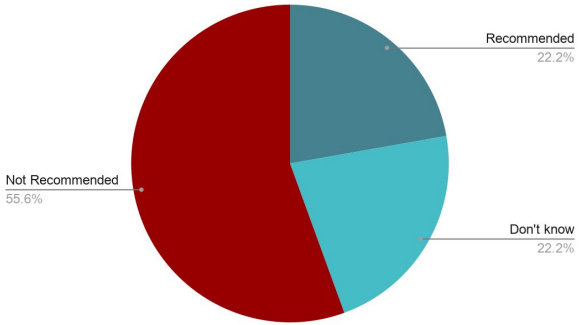
- it's not necessarily quantity but quality of social interactions that matter which is not necessarily measurable by the outcomes listed
- Unsure about frequency/duration of social interactions given that the ways to measure it don't necessarily seem to capture it, whereas duration/frequency of calls seems like a better way.
- Our earlier meetings suggested to me these are really compelling measures
- Clearly social connection is a critical outcome measure for mental health but not sure how well correlated these particular measures are with actual improvement given that social connection can be varied and personalized; it's more a perception that one is connected than actually the number of contacts one has
- I am not selecting frequency/duration of social interactions only because the digital ways to measure it don't seem to capture actual interaction (e.g., GPS locations does not equate to social interaction).
- Measures of social interaction derived from location, device use, and communication metadata are promising but currently lack sufficient specificity, interpretability, equity, and ethical clarity to warrant inclusion in a mental health digital measure set.
- … The potential confounding factor with frequency and duration of social interactions as well as physical proximity to others is that we are living in a more digital world, and more individuals are relying on online communities for their social interaction….

Measures for **Physical proximity to others**

Round 1



Round 2



Reported Reasoning for Inclusion:

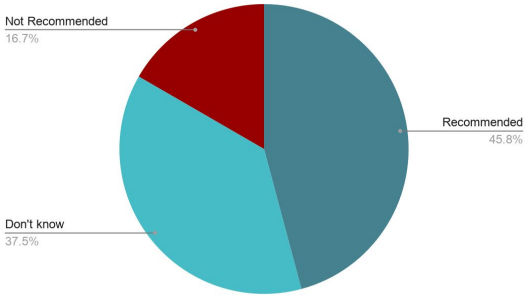
- NBDC reads as a high accessible and useful approach which I have been following
- Passively monitoring behavior is an excellent way to gain insight into mental health functioning.
- Interesting clinically + widely available measure
- I'd recommend these but with analyze data with caution. Controlling for other variables will be crucial here

Reported Considerations or Reasoning for Exclusion:

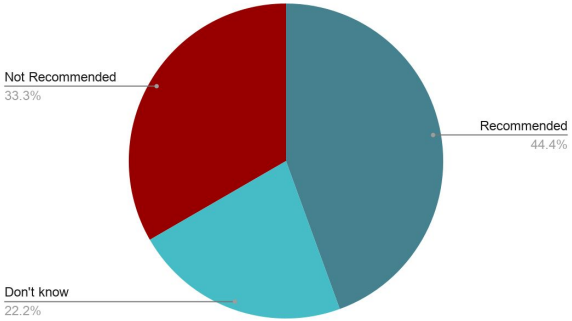
- Proximity to others I feel is not a ground truth because many do not have choice in this.
- What variability will one see with physical proximity and how is it necessary,,,?
- There are big privacy issues with content of speech collection, even in a research setting. Similar with Bluetooth.
- Physical proximity to others also does not necessarily capture social interaction in a meaningful way.
- Regarding bluetooth, it is unclear if this would be bluetooth in and outside of the home or only within the home. Some of these data are hard to interpret, as patients can be near many people but not experiencing quality social interactions.
- Physical proximity to others hard without context.
- Understanding social interaction can be beneficial... relying on devices for physical proximity might be too prone to errors.
- Measures of social interaction derived from location, device use, and communication metadata are promising but currently lack sufficient specificity, interpretability, equity, and ethical clarity to warrant inclusion in a mental health digital measure set.
- A balance between privacy and utility must be struck. We were involved in a bluetooth experiment - unfortunately the information gained by turning on bluetooth can also be misused by bad actors.

Measures for **Content & amount of speech & language**

Round 1



Round 2



Reported Reasoning for Inclusion:

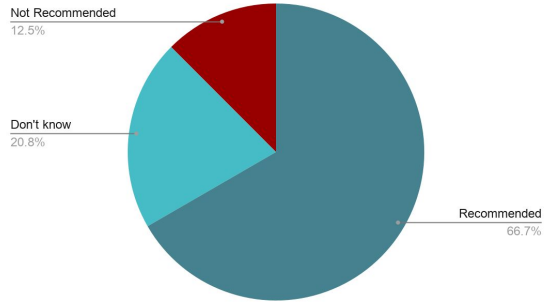
- Speech is a good marker of depression, mania and anxiety.
- that leaves speech and language (content and amount) as the best potential measurement
- I'd recommend these but with analyze data with caution. Controlling for other variables will be crucial here
- A key factor in mood
- Content and duration/frequency of calls and text have been seen across various mental health disorders...
- social interaction and social networks probably play a large role as well. again, not sure if they're core or if they can provide additional information

Reported Considerations or Reasoning for Exclusion:

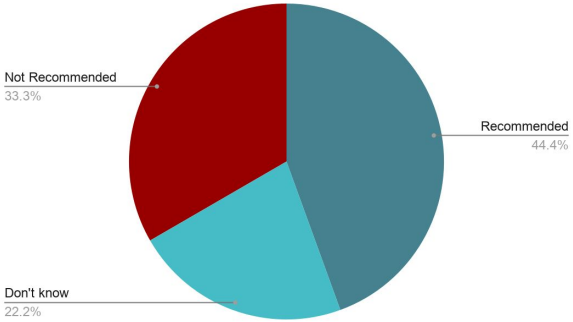
- If there was an effective measure of "Content and amount of speech and language (speech frequency and duration, and device usage; time spent around speech, topics on Twitter)", that was not extremely invasive, then I would be more interested since this address valence/aspects of quality.
- There are big privacy issues with content of speech collection, even in a research setting. Similar with Bluetooth.
- Same goes for content and amount of speech: unless done in a specific task, I think this is too privacy invasive.
- As for content, I would not focus exclusively on social media platforms, as that inherently excludes people who don't use them.
- I am not sure about content and amount of speech language especially as it relates to social media posts.
- Amount and content of speech has some associations but feels way too contextual and noisy to be a central feature? But not against be inclusive. But taking specific speech/text for analysis seems more standardized
- For content of speech and language, it is not clear to me if this is typed texted quantifiable using a passive mobile sensing app or if this would be a listening device to pick up on speech in the environment.

Measures for **Duration/frequency of calls or text**

Round 1



Round 2



Reported Reasoning for Inclusion:

- In addition, if we can collect information about frequency of text across all platforms as opposed to just phones, that might be more informative.
- We find links with cal time etc and by content I mean natural language processing not topics on twitter
- Content and duration/frequency of calls and text have been seen across various mental health disorders to aid in better characterizing mental health disorders. While there are hypotheses that the behavior does lead to a higher prevalence of mental health disorders, utilizing those measures in the process of measuring mental health could lead to insights/patterns that may not completely represent behaviors of individuals as they evolve in the digital age.
- Passive mobile sensors can also be used to detect call statistics and text messaging patterns, a real-world measure of social engagement. These measures have been associated with depression and manic symptoms clinical scales.

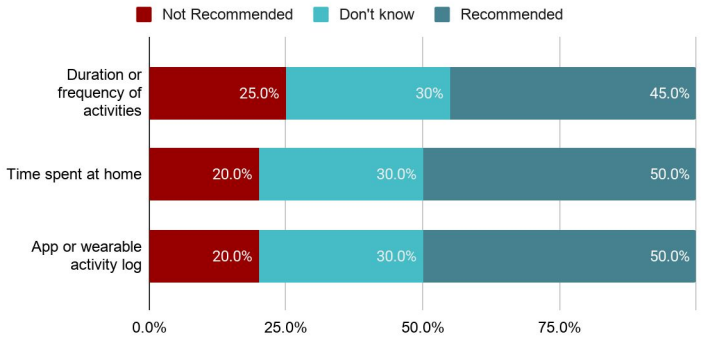
Reported Considerations or Reasoning for Exclusion:

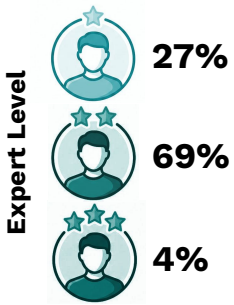
- It's hard for me to say how reliably these measures would be for mental health because the volume of calls/text could signal wellbeing or distress, or just changes in work demands or caregiving.
- Guess you could try to estimate from phone conversations whether they were ‘social’ or not, but haven't seen that done
- Our texting and calling data have small correlations with daily functioning
- I did not recommend duration/frequency of calls/texts because these interactions involve use of cell phones... and does not necessarily translate to meaningful social interaction.
- Duration/frequency of calls or text - not enough context captured...As for calls, there will need to be a distinction between people who use their phones for work and those who use them only socially.

Measures for **Personally meaningful activities**

DATACC by DiME

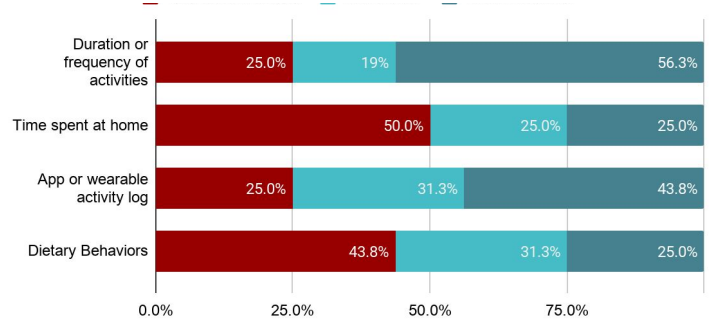
Round 1

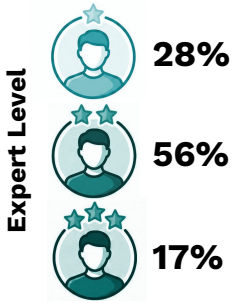




"Change in activities are the best indicators of worsening of health."

Round 2

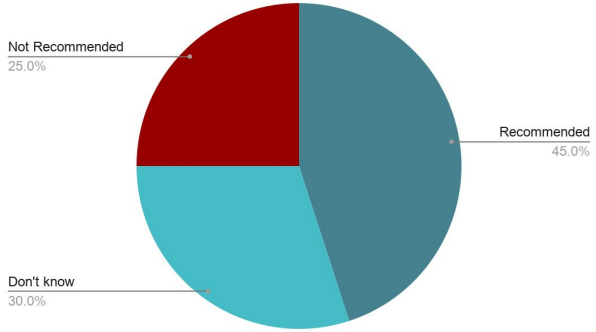




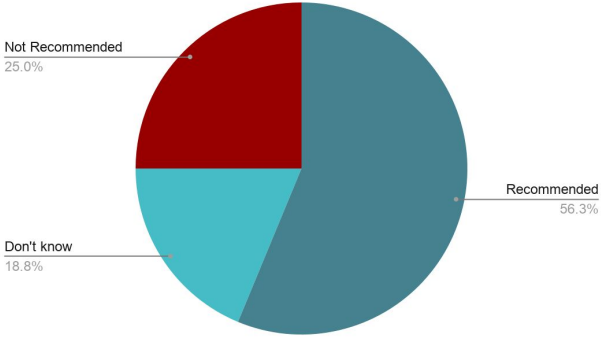
"Not recommended- privacy concerns and the subjective nature of the measures"

Measures for **Duration or frequency of activities**

Round 1



Round 2



Reported Reasoning for Inclusion:

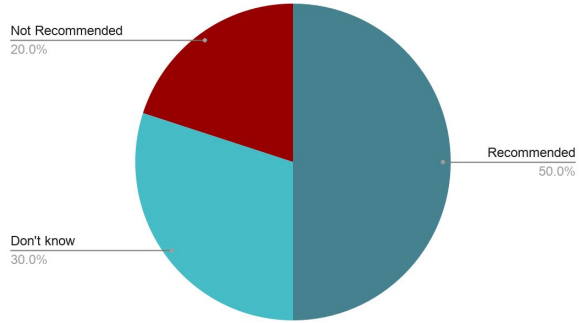
- As for duration and frequency of activities, I think behavioral activation is well-researched but the question is how to measure what is meaningful activities to a person.
- Behavioral activation is a very well studied technique.
- *All the three* above are important.
- We already capture this and it has been very valuable to our clinicians
- Duration or frequency of activities fits best.

Reported Considerations and/or Reasoning for Exclusion:

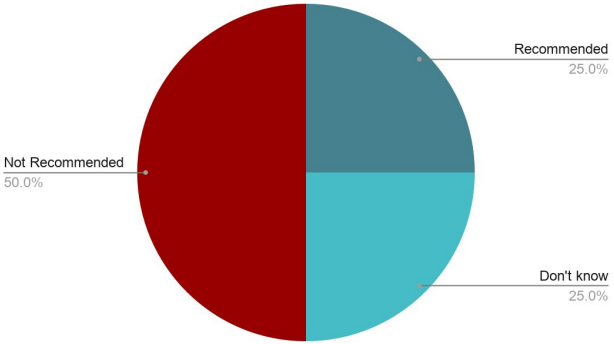
- similar to the last question, engagement in personally meaningful activities is more of a perception than an objective calculation
- I think we can only do this with questionnaires, as what is meaningful for one person... could be indistinguishable from unpleasant isolation for another person
- I did not select any of these measures as I have a hard time connecting how the defined measures actually measure engagement in personally meaningful activities.
- I would break down activity in a more granular way to include *dietary behaviors*
- These measures are promising but currently lack sufficient specificity, interpretability, equity, and ethical clarity to warrant inclusion in a mental health digital measure set.

Measures for **Time spent at home**

Round 1



Round 2



Reported Reasoning for Inclusion:

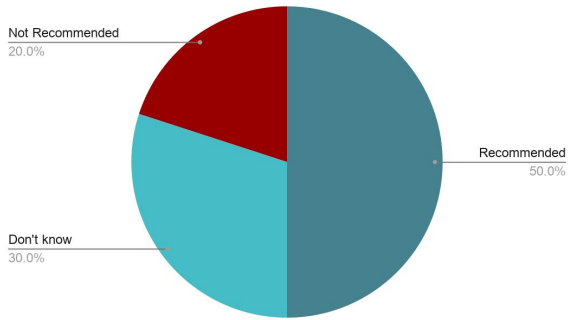
- I think time at home is the most privacy-preserving measure of outside interactions.
- Time spent at home could be useful only in combination with other measures (e.g. number of devices in proximity, sedentariness, sleep).
- Personally meaningful activities is so variable, but I think looking for changes in routine, like time at home, or not, is a good measure that doesn't infringe on privacy.

Reported Considerations or Reasoning for Exclusion:

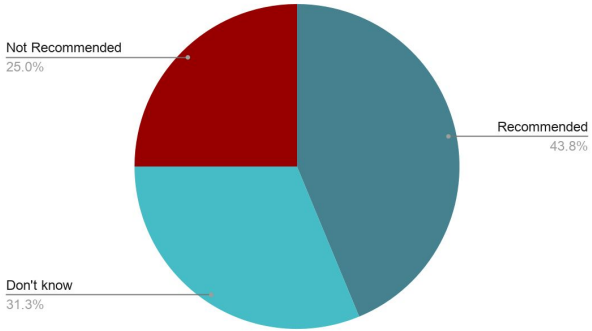
- We finds daily variability in areas from home travelled is strongly correlated with mental health; have not found links with time at home
- I also would exclude time spent at home because people can engage in meaningful activities at home.
- I'm unsure of time spent at home, i don't necessarily think that is indicative one way or another.
- Time spent at home is not as predictive as variability in location within our data"
- Once again with these measures, reliance on just GPS or time spent at home may not end up providing complete information on mental health disorders...
- Time spent at home I would say no, because a person can perform meaningful activity at home (e.g. knitting).
- These measures are promising but currently lack sufficient specificity, interpretability, equity, and ethical clarity...

Measures for **App or wearable activity log**

Round 1



Round 2



Reported Reasoning for Inclusion:

- I think the 'log' gives more context aware features than any one feature therein, phone use time within person may be good to detect things like 'doom scrolling'. our team has published on that.
- I don't feel that the other measures are tapping into meaningful correlates of mental health, aside from app or wearable log.
- There is a good amount of research specifically on social media use and mental health and this would feel like a "less invasive" and easy to use option.
- I think this can be done suitably privately.
- We already capture this and it has been very valuable to our clinicians
- I do believe there has been extensive research on app and phone usage, however, that has been insightful...

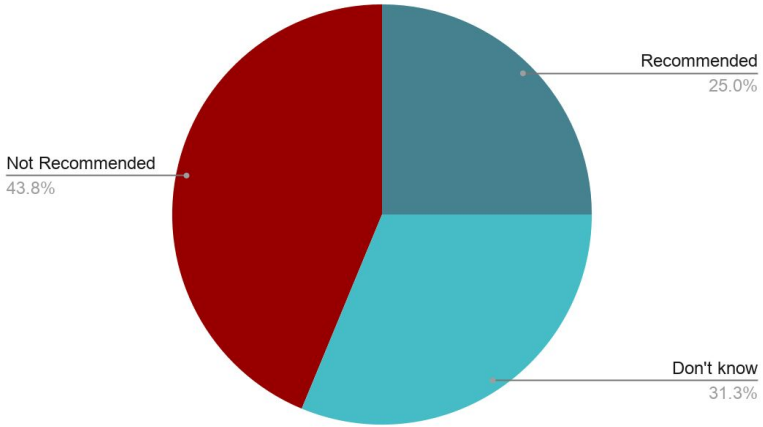
Reported Considerations or Reasoning for Exclusion:

- App log history we get and it takes quite a bit to make use of-but TBD how useful
- I would be interested in specifically social media app use given recent research...
- Maybe total phone use, but maybe it’s because I’m less familiar with this research but feels too noisy...
- My opinion is that it's too invasive, people won't like it
- App or wearable activity log I also would think hard to contextualize.
- These measures are promising but currently lack sufficient specificity, interpretability, equity, and ethical clarity...

Measures for **Dietary behaviors**

DATACC by DiME

Round 2



Reported Reasoning for Inclusion:

- Glycemic patterns are increasingly shown to be associated with depression subtypes and anxiety symptoms. I think it would be beneficial to fully understand the big picture of how metabolic response affects mental health and its interrelationship with sleep, cortisol, and inflammation.

Reported Considerations or Reasoning for Exclusion:

- I would not recommend dietary behaviors given often low compliance/reporting bias in food diaries and transdiagnostic relevance.
- I would shy away from any tracking of this related to someone having to log food (this is associated with disordered eating habits and patterns).
- Dietary behaviors I'm always suspect of- I just don't think the science of CGM is there quite yet and self-reports are horrible.

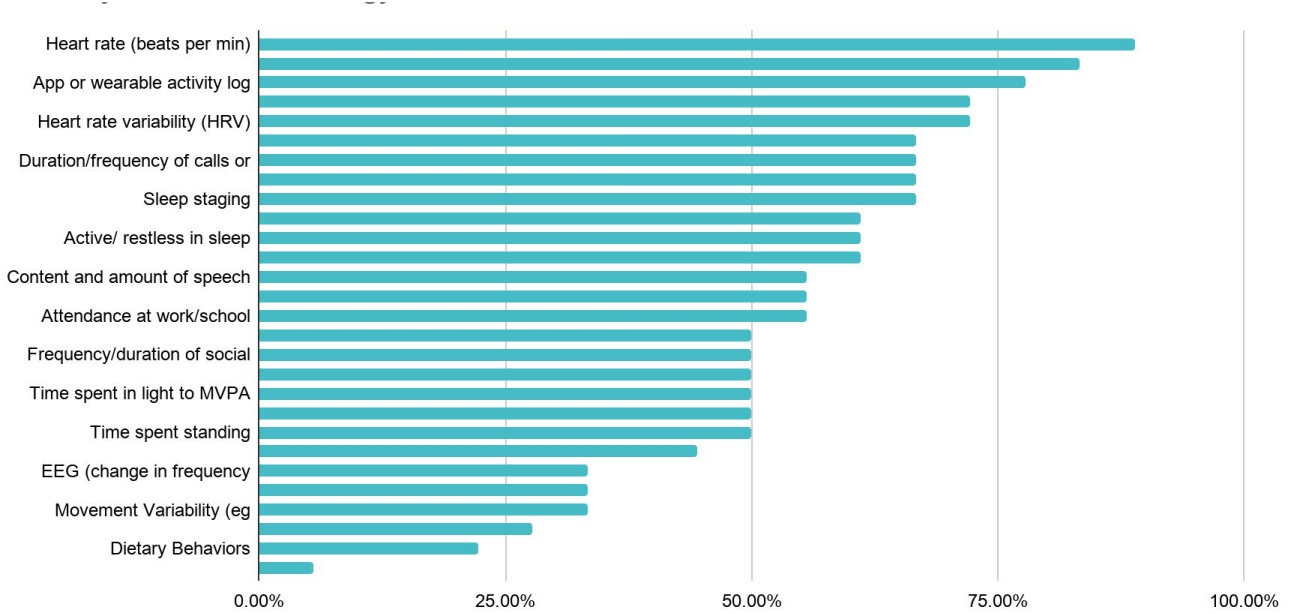
Digitally-Derivable Measures: Round 2

DATACC by DiMe

Which of the previously identified measures do you agree can be digitally-derived with currently available technology?

In agreement with Round 1, the majority of measures were supported for digital collection.

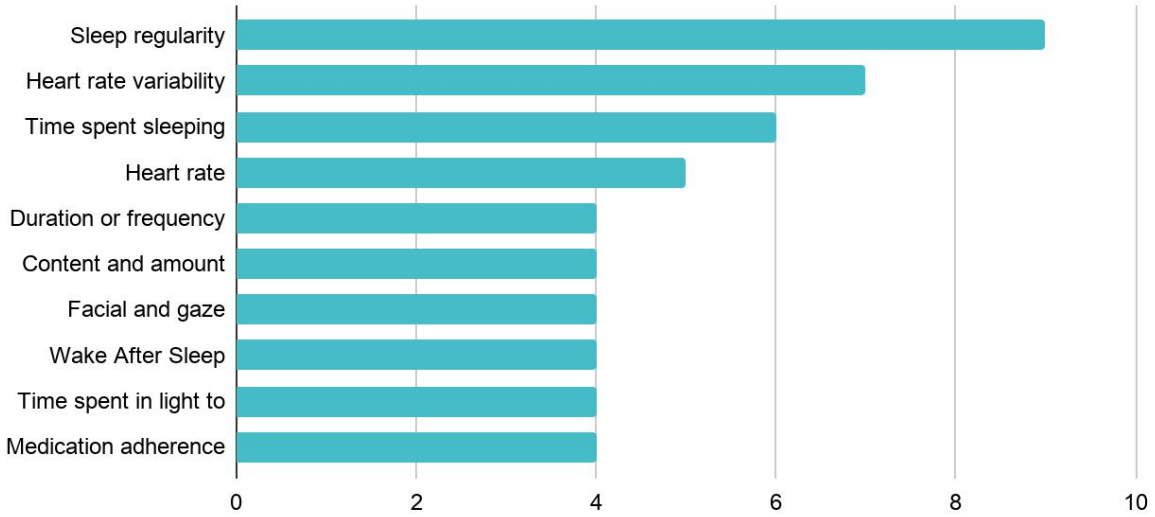
Additional measures dropping below 50% include **facial valence**, **movement variability**, **postural stability on standing**, **and dietary behaviors**.



'Minimally Necessary' Measures: Round 2

Of the previously identified measures, which measures, if any, are minimally necessary or required for inclusion into a digital measures set for CMHD?

'Minimally Necessary': Top 10 (n=18)



When asked to select the essentials, experts aligned on:

- Sleep regularity
- Heart rate variability

- Remaining measures fall below 50% in recommendation to the digital measures set for CMHD.

Recommendations for Measures of **Common Mental Health Disorders**

Concepts of Interest	Measures which proceed to next round *(informed by delphi results & debrief workshop selection decisions)*	Measures where there is not enough evidence to make a recommendation
Routine Adherence	Variability in daily routines Cognitive function (exploratory) Sleep regularity	Attendance at work/school Medication Adherence
Movement	Step Count Daily average minutes of physical activity per time Time spent in MVPA (emerging)	Movement Variability Time sedentary Time spent standing Postural stability on standing
Sleep	Sleep onset latency Sleep efficiency WASO (Wake After Sleep Onset) (emerging)	Active/ restless in sleep window Time spent sleeping Sleep staging
Frequency, intensity, & Change in Emotions	Change in speech and language Heart rate variability (HRV) measures EDA (emerging) Facial and gaze features (emerging)	Heart Rate EEG Facial valence
Social Interactions		Frequency/duration of social interactions Content and amount of speech and language Duration/frequency of calls or text Physical proximity to others
Personally Meaningful Activities		Duration or frequency of activities App or wearable activity log Time spent at home Dietary Behaviors

Recommendations for Measures of **Common Mental Health Disorders**

Concepts of Interest	Measures which proceed to next round *(informed by delphi results & debrief workshop selection decisions)*
Routine Adherence	Variability in daily routines Cognitive function Sleep regularity
Movement	Step Count Daily average minutes of physical activity per time Time spent in MVPA
Sleep	Sleep onset latency Sleep efficiency WASO (Wake After Sleep Onset)
Frequency, intensity, & Change in Emotions	Change in speech and language Heart rate variability (HRV) measures EDA Facial and gaze features
Social Interactions	
Personally Meaningful Activities	

Next steps to identifying the digital measures of common mental health disorders

- Determine the level of evidence (verification, usability validation analytical validation, and clinical validation) associated with each digital meaus, measured in real world settings.
- Determine digital measures which are about to be captured by emerging digital health technologies on the market.
- Identify a narrow digital measures set based on the “readiness” for their use in clinical research and care.

Recommendations for Measures of **Common Mental Health Disorders**

Concepts of Interest	Measures where ther is not enough evidence to make a recommendation
Routine Adherence	Attendance at work/school Medication Adherence
Movement	Movement Variability Time sedentary Time spent standing Postural stability on standing
Sleep	Active/ restless in sleep window Time spent sleeping Sleep staging
Frequency, intensity, & Change in Emotions	Heart Rate EEG Facial valence
Social Interactions	Frequency/duration of social interactions Content and amount of speech and language Duration/frequency of calls or text Physical proximity to others
Personally Meaningful Activities	Duration or frequency of activities App or wearable activity log Time spent at home Dietary Behaviors

- Measures where <75% of delphi survey experts did not select their inclusion in the set of digital measures for CMHD and an expert stakeholder group confirmed exclusion during the debrief workshop.
- There is not enough evidence to support these measure as digital measures for CMHD; however, evidence may support excluded measures for specific disorders.
- Recommendations for more innovation and research to support measure related to these concepts of interest, particularly for Social Interactions and Personally Meaningful Activities.

Modified Delphi Approach to Identify a Consensus List of Digital Measures

The set of digital measures for common mental health disorders were narrowed down based on their readiness for widespread adoption in clinical research and care. To inform selection, we reviewed:

- ***Maturity of technologies and measures**, including v**erification, analytical validation, and clinical validation studies** conducted for measurement in real-world setting.*
- *Digital measures of MH being used in **clinical trials as endpoints**.*
- ***Narrative review** of digital measurement for common mental health disorders, extending the previous global research report.*

Based on this evidence, the set of digital measures for common mental health disorders relate to: (1) variability in daily routines (2) cognitive function (3) physical activity (4) sleep (5) speech & language (6) autonomic function and (7) facial and gaze features

Thank you

Samantha McClenahan | Samantha.McClenahan@dimesociety.org

Candice Taguibao | Candice@dimesociety.org

Danielle Stefko | danielle@dimesociety.org

DATACC

Digital Health Measurement Collaborative Community

by DiME

www.datacc.dimesociety.org

www.linkedin.com/company/dime-society

https://dimesociety.slack.com

Survey Completion Overview

Round 1

Avg. Time to Complete
24m

Completed
26

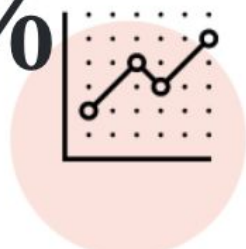
Completion Rate
100.00%

Round 2

Avg. Time to Complete
17m

Completed
18

Completion Rate
100.00%



A Technical and Clinical Expert Panel

Survey Respondents (N=26) weighted toward measurement feasibility, evidence, and digital execution for CMHD.

	Round 1	Round 2
Digital health & digital health technologies	20	9
Measurement science & endpoints	13	9
Clinical practice	12	8
Data science, informatics & AI/ML	8	7
Clinical science & operations	7	6
Regulatory & policy	6	4
Patient advocacy	5	3
Lived experience	5	3
Common Mental Health Disorders R&D	4	2

Key Insight: The panel is highly specialized in the execution of digital measures, with strong representation in product development and endpoint validation.

Digital Measures Set: Analysis

- **Descriptive statistics**
  - Data analysis:
    - 1. Weighted - all participants
      - Knowledge: Limited=x1, Moderate=x2, Extensive=x3
    - **2. Weighted [Shared data]- subgroup of participants with Moderate & Extensive knowledge**
      - **Knowledge: Limited=x0, Moderate=x2, Extensive=x3**
  - Evaluating amount of:
    - % of support for a respective answer (yes-es)
    - % of ambiguity (don't know-s)
    - % of opposition (no-s)

- **Qualitative inputs**
  - Statistical insights corroborated with qualitative inputs

Digital Measures Set: Analysis

- **Selecting the digital measures set:**
  - If a measure reached **>75% consensus** in **ALL** participants -> **include**
  - If a measure didn't reach **>75% consensus** in all participants, but it did so in participants with **MODERATE/EXTENSIVE** expertise -> **include**
  - If a measure was recommended in Round 1 -> assess in **Round 2**
  - If a measure reached 25% - 74.9% consensus -> Reassess in **Round 2**
  - If a measure reached **<25% consensus** -> exclude from **Round 2**

  - If a question didn't have **at least 8 experts** with MODERATE/EXTENSIVE expertise -> a decision that we **can't make a recommendation yet**
    - 8 experts is based on recommendations from this source: *Delphi panel sizes should ideally be between 8 to 23 participants.*