Mental Health Across the Care Continuum: A Review and Technology-Powered Reference Standard

Matthew Vowels[1,2,*] (Ph.D Eng., Ph.D Appl. Math) **Hank Capps**[1,4] (M.D.) **Kathryn Erickson-Ridout**[6] (MD, Ph.D) **Laura Vowels**[3] (Ph.D) **Ish Bhalla**[5] (MD) **Carol Alter**[7] (MD) **Philip D. Harvey**[8] (Ph.D) **Charles Nemeroff**[7] (MD, Ph.D)

[1] *Kivira Health, NY, USA*
[2] *The Sense, Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland*
[3] *University of Roehampton, London, United Kingdom*
[4] *Seny Advisory, Kivira Health*
[5] *Yale University, Duke University School of Medicine, USA*
[6] *Kaiser Permanente Division of Research, USA*
[7] *Dell Medical School, University of Texas at Austin, Austin, TX, USA*
[8] *Miller School of Medicine, University of Miami, Miami, FL, USA*

[*] Correspondence: matthew.vowels@chuv.ch

Background: Mental health systems globally face persistent structural failures in identification, assessment, treatment access, continuity of care, and evidence integration. Most individuals requiring care are never identified or are identified years after disorder onset, while diagnostic accuracy and treatment pathways remain inconsistent across settings. Simultaneously, artificial intelligence and digital mental health technologies are rapidly proliferating without standardized governance, validation, interoperability, or evidence-traceability frameworks.

Objective: To propose a vendor-neutral integrated mental health assessment infrastructure specification defining the core capabilities technology-enabled mental health systems should support across the continuum of care.

Methods: This work synthesizes evidence and operational requirements across 21 symptom and disorder domains, including depression, anxiety, psychotic-spectrum disorders, suicidality, substance use, eating disorders, neurodevelopmental conditions, and interpersonal functioning. The review additionally evaluates 17 assessment strategies across more than 30 implementation and validity dimensions, treatment pathways, reimbursement structures, and 16 care settings spanning primary care, specialty psychiatry, emergency medicine, inpatient care, telehealth, and correctional systems.

Results: We define a reference standard for integrated mental health assessment systems encompassing: (1) structured multi-domain assessment with dynamically administered questionnaires; (2) longitudinal measurement-based reassessment and symptom tracking; (3) real-time risk detection and safety escalation; (4) evidence-based clinical decision support with transparent and traceable evidence sources; (5) navigation of care delivery and continuity across transitions; (6) Electronic Health Record interoperability with auditability; and (7) a governed evidence layer supporting continuous re-validation of instruments, guidelines, and predictive systems using real-world deployment data.

The proposed framework integrates digital self-report, structured assessment, clinical guidance, and workflow-aware implementation while addressing constraints relating to scalability, patient safety, reimbursement, operational logistics, and clinical validity. Conclusions: A standardized, interoperable infrastructure for mental health assessment may improve systematic identification, clinical decision-making, continuity of care, and deployment governance for digital and AI-enabled systems. The proposed specification provides a scalable foundation for integrating evidence-based assessment and measurement-based care across heterogeneous clinical environments while enabling continuous evaluation and updating of the underlying evidence base.

Keywords: mental health assessment; measurement-based care; clinical decision support; artificial intelligence; digital mental health; psychiatric diagnosis; interoperability; healthcare infrastructure; evidence governance; risk detection.

Contents

Abbreviations and Acronyms **3**

1 Introduction **5**

2 The state of mental health care **5**

2.1 The core problem . . . 5
2.2 The scope of mental health conditions and presentations . . . 7
2.3 The treatment landscape . . . 7
2.4 Measurement modalities and delivery formats . . . 8
2.5 Care settings and populations . . . 8
2.6 Financial structures, reimbursement, and economic feasibility . . . 13
2.7 Strategic prioritisation . . . 14
2.7.1 Where impact is greatest and most achievable . . . 14
2.7.2 Where the challenges are most complex . . . 15
2.7.3 Treatment-resistant and complex presentations . . . 15
2.7.4 Conditions where self-report is insufficient . . . 15
2.7.5 Conditions where the treatment landscape itself is contested . . . 15
2.7.6 Emerging measurement modalities . . . 15
2.7.7 Populations underserved by existing instruments . . . 15
2.7.8 Personalised treatment sequencing . . . 20

3 The proposed technology-powered reference standard **20**

3.1 Design constraints and feasibility requirements . . . 20
3.2 The reference model . . . 22
3.3 Safety, privacy, and equity requirements . . . 22
3.4 Regulatory positioning, implementation, and governance . . . 22
3.5 Alternative approaches and counterpositions . . . 23

4 Conclusion and call to action **24**

About the Authors **26**

References **28**

A Clinical Scope of Mental Health Conditions and Presentations **40**

B Treatment Landscape **42**

C Measurement Modalities **43**

D Care Settings **47**

E Financial Structures and Economic Considerations **49**

Abbreviations and Acronyms

ABA	applied behaviour analysis
ACT	acceptance and commitment therapy
ADHD	attention-deficit/hyperactivity disorder
AN	anorexia nervosa
AQ-10	Autism Spectrum Quotient (10-item)
ASQ	Ask Suicide-Screening Questions
ASRM	Altman Self-Rating Mania Scale
ASRS v1.1	Adult ADHD Self-Report Scale (version 1.1)
AUD	alcohol use disorder
AUDIT	Alcohol Use Disorders Identification Test
BA	behavioural activation
BDD	body dysmorphic disorder
BED	binge eating disorder
BN	bulimia nervosa
BPD	borderline personality disorder
C-PTSD	complex post-traumatic stress disorder
CAMS	Collaborative Assessment and Management of Suicidality
CAPE	Community Assessment of Psychic Experiences
CBT	cognitive behavioural therapy
CBT-E	enhanced cognitive behavioural therapy
CBT-I	cognitive behavioural therapy for insomnia
CBTp	cognitive behavioural therapy for psychosis
CDS	clinical decision support
CM	contingency management
CoCM	Collaborative Care Model
COI	conflict of interest
CPT	cognitive processing therapy
C-SSRS	Columbia-Suicide Severity Rating Scale
DAST	Drug Abuse Screening Test
DBS	deep brain stimulation
DBT	dialectical behaviour therapy
dTMS	deep transcranial magnetic stimulation
DTx	digital therapeutic
EAP	employee assistance programme
EDA	electrodermal activity
ED	emergency department
EDys	erectile dysfunction
EEG	electroencephalography
EHR	electronic health record
EMA	ecological momentary assessment
EMDR	eye movement desensitisation and reprocessing
ERP	exposure and response prevention
ERPs	event-related potentials
eTNS	external trigeminal nerve stimulation
FBT	family-based treatment
FGA	first-generation antipsychotic
FHIR	Fast Healthcare Interoperability Resources
FSFI	Female Sexual Function Index
GAD	generalised anxiety disorder
GAD-7	Generalized Anxiety Disorder 7-item scale
HAM-A	Hamilton Anxiety Rating Scale
HRV	heart rate variability
ICD-11	International Classification of Diseases (11th Revision)
IIEF-5	International Index of Erectile Function (5-item)
IPT	interpersonal psychotherapy
IPSRT	interpersonal and social rhythm therapy
ISI	Insomnia Severity Index
LAI	long-acting injectable
MADRS	Montgomery–Åsberg Depression Rating Scale
MAOI	monoamine oxidase inhibitor
MANTRA	Maudsley Anorexia Nervosa Treatment for Adults
MAT	medication-assisted treatment
MBC	measurement-based care
MBCT	mindfulness-based cognitive therapy
MBSR	mindfulness-based stress reduction
MI	motivational interviewing
MINI	Mini International Neuropsychiatric Interview
MMSE	Mini-Mental State Examination
MoCA	Montreal Cognitive Assessment

MRI	magnetic resonance imaging
MST	magnetic seizure therapy
NICE	National Institute for Health and Care Excellence
NLP	natural language processing
OCD	obsessive-compulsive disorder
OSFED	other specified feeding or eating disorder
OUD	opioid use disorder
PE	prolonged exposure
PDE5	phosphodiesterase type 5
PDSS-SR	Panic Disorder Severity Scale (self-report)
PET	positron emission tomography
PHQ-2	Patient Health Questionnaire-2
PHQ-9	Patient Health Questionnaire-9
PPV	positive predictive value
PTSD	post-traumatic stress disorder
qEEG	quantitative electroencephalography
RCT	randomised controlled trial
RDoC	Research Domain Criteria
rTMS	repetitive transcranial magnetic stimulation
SCID-5	Structured Clinical Interview for DSM-5
SCOFF	Sick, Control, One stone, Fat, Food questionnaire
SD	sexual dysfunction
SGA	second-generation antipsychotic
SI	suicidal ideation
SNRI	serotonin–norepinephrine reuptake inhibitor
SSRI	selective serotonin reuptake inhibitor
SUD	substance use disorder
TBI	traumatic brain injury
TCA	tricyclic antidepressant
tDCS	transcranial direct current stimulation
TF-CBT	trauma-focused cognitive behavioural therapy
TMS	transcranial magnetic stimulation
TMT	Trail Making Test
TRD	treatment-resistant depression
VNS	vagus nerve stimulation
WCST	Wisconsin Card Sorting Test
Y-BOCS	Yale–Brown Obsessive Compulsive Scale

1 Introduction

This work proposes a reference standard for technology-powered mental health care spanning screening, structured assessment and generation of diagnoses, prevention, longitudinal monitoring, clinical decision support in evidence-based practice, and integration and care coordination across the care continuum. The proposal has two aims. First, it establishes why such a standard is needed, by documenting the structural limitations of current mental health care delivery, reviewing the methodological concerns in the evidence base on which clinical decisions depend, and commenting on the rapid proliferation of digital mental health technologies without shared frameworks for evaluation or governance. Second, it defines a vendor-neutral reference model, specifying capabilities that technology-powered mental health systems in this class should support, spanning multi-domain assessment, measurement-based care, risk detection, treatment decision support, care navigation, and reporting. We do not describe or advocate for a specific product, instead we describe a class of systems and the evidentiary standards that should apply to them.

The reference standard proposed in this work emerges as a result of a review of mental health across care settings, clinical conditions, measurement modalities, treatment complexity, workforce constraints, scalability requirements, ethical considerations, and regulatory structures. When these dimensions are evaluated simultaneously, a specific class of system emerges as the baseline infrastructure the field needs: transparent, longitudinal, structured assessment systems that seamlessly integrate validated instruments with evidence-based clinical decision support into clinical workflows, care navigation, and governed evidence management.

Section 2 presents a review of the current state of mental health care, covering the structural problems in care delivery, the scope of conditions and most common conditions treated, the available measurement modalities, the care settings and populations served, the financial and reimbursement structures that constrain service delivery, and a strategic prioritisation analysis that identifies where the intersection of impact and feasibility is greatest. Section 3 presents the proposed reference standard, including the reference model, evaluation framework, safety and privacy requirements, equity considerations, implementation pathways, regulatory alignment, and governance of the standard. Section 4 provides a conclusion and call to action. The central claim of this work is that scalable, structured, longitudinal assessment grounded in validated instruments and governed evidence is the missing infrastructure layer required to improve mental health care and patient outcomes at scale.

2 The state of mental health care

2.1 The core problem

Mental health and neurological disorders account for approximately 10.4% of disability-adjusted life years worldwide (Whiteford et al., 2013). Despite substantial advances in treatment efficacy, the systems designed to detect and treat these conditions face persistent structural limitations, which include substantial lack of access to trained and experienced practitioners (Katayama et al., 2023; Thornicroft et al., 2017). Current digital mental health tools are typically fragmented, condition-specific, weakly integrated into care pathways, and inconsistently developed or validated therefore limiting their clinical utility and scalability. Six converging challenges in mental health service delivery motivate the present document:

1. Many cases of serious mental illness are never identified, and those that are identified are identified late. Fewer than half of individuals with diagnosable mental disorders receive any professional care (Kohn et al., 2004; Thornicroft et al., 2017). Among those who do seek treatment, delays of 6–23 years between symptom onset and first clinical contact are common, depending on the disorder (Wang et al., 2005). A substantial proportion of cases are never detected at all - not because assessment fails, but because affected individuals never present for care, particularly in systems with limited outreach, inadequate community-based services, or structural barriers to access. For those who do reach clinical settings, traditional case-finding often depends on unstructured interviews during brief primary care visits - a model sensitive to clinician training, time constraints, the salience of presenting complaints, and the patient's likelihood to seek accessible care. Detection rates under these conditions are known to be suboptimal (Carey et al., 2014; Vermani et al., 2011).

2. When cases are identified, they are frequently misclassified. Detection of depression and anxiety-related disorders in primary care is often limited, with unassisted GP detection of depression showing low sensitivity and studies of mood and anxiety disorders reporting high rates of missed or incorrect recognition (Carey et al., 2014; Vermani et al., 2011). Initial diagnoses in mental health settings can also change substantially on further assessment: rates of diagnostic revision for schizophrenia have been reported in the range of 36-51% (Bradford et al., 2024); patient surveys of bipolar disorder report that 69-73% of respondents had previously received an incorrect diagnosis (Hirschfeld et al., 2003; Lish et al., 1994); and schizoaffective disorder has been reported as a particularly unstable or frequently misclassified diagnosis in some settings (Bradford et al., 2024). Diagnostic clarity is especially challenged where anxiety, reported hallucinations, comorbidity, or

overlapping symptom profiles complicate differential diagnosis (Coulter et al., 2019). Low or inconsistent use of validated screening and assessment instruments may further contribute to variability in recognition and management, particularly where clinicians differ in training, familiarity, or confidence with these tools (Waheed et al., 2024).

3. Assessment is typically a pretreatment event rather than a continuous process. Current practice treats screening as a snapshot rather than a trajectory. A single PHQ-9 score during a 15-minute primary care visit cannot capture the longitudinal dynamics of symptom onset, progression, treatment response, and relapse. In the absence of measurement-based care, detection of treatment non-response is delayed and recovery management proceeds without systematic outcome data. Current practice relies on snapshot screening rather than longitudinal measurement, despite evidence supporting measurement-based care approaches that incorporate repeated assessment over time (Fortney et al., 2017).

4. The pathway from assessment to appropriate treatment is inconsistently supported. Even when a positive screen is obtained, the availability of and referral to evidence-based treatment varies considerably across settings. Care navigation that matches patients to the appropriate modality (psychotherapy, pharmacotherapy, neuromodulation, digital therapeutics, peer support, lifestyle intervention), provider, and intensity level is not systematically supported in most care environments, despite evidence that stepped and matched care models can improve response and remission rates compared with usual care (van Straten et al., 2015). Furthermore, referral pathways vary in structure, scheduled therapy sessions are often widely dispersed, and between-session support is limited. These factors contribute to discontinuities along the care continuum, particularly at the transitions between assessment and treatment initiation, and between active treatment and sustained recovery management; systematic reviews of initial appointment non-attendance and early dropout document substantial losses at exactly these transition points (Sweetman et al., 2021). Repeated misclassification, poorly matched treatment, or unsupported referrals may also erode trust in care, reducing adherence and increasing the likelihood that patients disengage until symptoms escalate to crisis or emergency presentation.

5. Emerging technology is advancing faster than the standards needed to govern it (Chan et al., 2025; Lecomte et al., 2020; Torous et al., 2018). Large language models, adaptive assessment engines, and AI-driven clinical decision support tools are entering mental health care without consensus on what constitutes safe, effective, and equitable deployment (Papagiannidis et al., 2025; World Health Organization, 2021a). The regulatory landscape is fragmented across jurisdictions and some jurisdictions in the US have already placed restrictions on digital therapy assistants that render them challenging to deploy. No shared framework exists for evaluating these systems in proportion to the clinical claims they make and the risks they may introduce. Without a reference standard, the result is likely to be a heterogeneous collection of inconsistently evaluated tools with variable safety and efficacy profiles.

6. The evidence base underpinning clinical decisions has known limitations that must be actively managed. The challenges described above are compounded by well-documented methodological concerns in the scientific literature that underpins mental health assessment and treatment. Publication bias, selective outcome reporting, small and unrepresentative samples, heterogeneous methods, overfitting of diagnostic cutoffs to derivation samples, and insufficient replication are pervasive (Flake & Fried, 2020; Haven & Ioannidis, 2026; Ioannidis, 2005; Turner et al., 2008; Vowels, 2023). For screening instruments specifically, many validation studies derive optimal cutoffs on the same sample used to estimate accuracy, head-to-head comparisons within domains are rare, reference standards vary across studies, and measurement invariance testing across demographic subgroups remains incomplete for most measures. Treatment guidelines are similarly affected by these concerns. The literature from which they are derived reflects incentive structures that reward novelty over replication, statistical significance over clinical significance, and publication volume over methodological rigour. Head-to-head comparisons between bona fide psychotherapy modalities consistently yield small or non-significant differences, and direct comparisons between psychotherapy and pharmacotherapy show similarly modest differentiation - raising fundamental questions about the evidential basis for recommending one specific treatment over another for a given patient. AI systems trained or prompted on this literature, and technology-powered systems that encode these guidelines, inherit whatever biases and inaccuracies the underlying evidence contains.

This is not, however, an argument against the use of validated instruments and evidence-based guidelines. On the contrary: standardized, psychometrically evaluated screening tools represent one of the most defensible starting points, outperforming unstructured clinical judgement in terms of consistency, reproducibility and scalability (Ægisdóttir et al., 2006; Grove et al., 2000). Many instruments have been subjected to more scrutiny than the vast majority of clinical assessment practices they would augment or replace. The alternative - continued reliance on unstandardized, clinician-variable assessment - is not a safer option. At the same time, structured instruments and clinical judgment serve com-

plementary roles: instruments provide reproducible, population-level signal, while clinicians interpret, contextualise, and act on that signal in light of individual patient circumstances. Neither alone is sufficient; improved outcomes depend on their integration within a structured, transparent, and iterative assessment process.

2.2 The scope of mental health conditions and presentations

A detailed discussion of conditions, prevalence, and burden is provided in Appendix A. For organisational clarity, we adopt the DSM-5-TR (American Psychiatric Association, 2022) as the primary reference framework, given its continued use in clinical practice, research, reimbursement systems, and regulatory contexts. At the same time, the limitations of categorical diagnosis are well established: diagnostic boundaries are often porous, comorbidity is pervasive, reliability is uneven across conditions, and many clinically significant presentations do not fit neatly within a single discrete category. Frameworks such as the ICD-11 (World Health Organization, 2022a), RDoC (Insel et al., 2010), and HiTOP (Kotov et al., 2017) highlight the need for approaches that can accommodate dimensionality, overlap, and heterogeneity rather than assuming that clinical reality is well captured by rigid categories alone.

Accordingly, the scope of this described reference standard extends beyond formal diagnoses. It includes high-prevalence conditions with many putative subtypes such as depressive disorders, generalised anxiety disorder, panic disorder, social anxiety disorder, PTSD, and OCD; severe mental illness including bipolar and psychotic-spectrum disorders; neurodevelopmental conditions such as ADHD and autism; alcohol and substance use disorders; feeding and eating disorders; insomnia and other sleep-related problems; sexual dysfunction; suicidality and self-harm; subjective cognitive decline; personality pathology, particularly borderline personality disorder; adjustment disorders, burnout, stress, emotional dysregulation, body dysmorphic disorder, specific phobias, and symptom-level or subthreshold presentations that may still require monitoring or intervention. It also includes relational and interpersonal functioning, family dysfunction, resilience, attachment-related difficulties, loneliness, and socio-economic status which are not DSM-5 diagnostic categories in themselves but are clinically significant drivers and moderators of outcomes across many disorders.

The central implication is that a viable assessment standard cannot be built around one disorder at a time. Mental health presentations are multi-domain, transdiagnostic, and often developmentally and contextually shaped, with high rates of within- and cross-spectrum comorbidity documented in large epidemiological surveys (Kessler et al., 2005a). A condition-specific approach might reproduce the fragmentation that already characterises much of mental health care, where patients are screened narrowly despite overlapping syndromes, mixed risk states, and functionally important symptoms outside a single diagnostic category.

At the same time, current clinical workflows, reimbursement structures, and regulatory frameworks remain largely organised around disorder-based classification systems such as DSM-5-TR. Accordingly, the reference standard is designed to operate pragmatically within these constraints: in the near term, this involves the structured combination of validated, disorder-specific instruments integrated through shared administration and contextual decision logic. Over time, as large-scale longitudinal data accumulates and validation frameworks mature, this approach creates a pathway toward more unified, transdiagnostic or symptom-level assessment models that better reflect underlying clinical reality while remaining actionable in practice – an evolution paralleled in research frameworks such as the Hierarchical Taxonomy of Psychopathology, which organises psychopathology along empirically derived dimensions rather than discrete categories (Kotov et al., 2017). The full clinical scope and epidemiological context are detailed in Appendix A.

2.3 The treatment landscape

A standard for assessment must be designed in light of the clinical decisions that follow from it. The treatment landscape is described in Appendix B and summarised in Table 1; this section highlights the implications most relevant to the design of the standard. Mental health treatment spans pharmacotherapy, psychotherapy, neuromodulation, digital therapeutics, and lifestyle or adjunctive interventions. Pharmacological treatment includes SSRIs, SNRIs, mood stabilisers, benzodiazepines, antipsychotics, stimulant and non-stimulant ADHD medications, and medications for substance use disorders. Psychotherapeutic care includes CBT, DBT, trauma-focused therapies such as EMDR, CPT, and prolonged exposure, third-wave approaches such as ACT and MBCT, interpersonal psychotherapy, psychodynamic approaches, and family or systemic therapies. Other parts of the treatment landscape include ECT, rTMS, ketamine and esketamine, emerging psychedelic-assisted therapies, digital therapeutics, exercise, sleep-focused interventions, diet, peer support, and social prescribing.

Across these modalities, the evidence base is heterogeneous in both quality and maturity, and is affected by the same metascientific problems identified in this document, including publication bias, problematic comparators (such as poorly blinded placebo), selective reporting, researcher allegiance effects, endpoint switching, and conflicts of interest; an influential analysis of FDA-registered antidepressant trials, for example, found

that selective publication inflated apparent effect sizes by approximately 32% across the drug class (Turner et al., 2008). In many areas of mental health, average treatment effects are modest, differences between commonly used interventions are often small, and combined treatment may outperform monotherapy depending on the condition, severity, and stage of care, as illustrated by network meta-analyses showing relatively narrow efficacy and acceptability differences across commonly used antidepressants (Cipriani et al., 2018). In practice, treatment choice is shaped not only by efficacy estimates but also by side effects, contraindications, drug interactions, comorbidity, patient preference, cost, availability, workforce constraints, and sequencing after prior non-response. Misclassification also creates direct patient-safety risks: unnecessary or poorly matched pharmacological treatment can expose patients to avoidable adverse effects, including metabolic, neurological, sexual, and weight-related harms, with particular concern where complex psychotropic prescribing occurs without specialist behavioral health input.

The design implication is that technology-powered systems should not present treatment as a deterministic output of assessment. Their appropriate role is to support structured, evidence-informed clinical reasoning by presenting options, trade-offs, and uncertainty in a transparent way, and to facilitate shared decision-making between clinician and patient through clear communication of risks, benefits, and alternatives. This requires systems to incorporate evidence grading, provenance, and ongoing re-validation rather than treating guidelines as fixed or exhaustive rules. In other words, **the assessment standard must be compatible with a treatment landscape in which the evidence is useful but incomplete, and in which treatment selection always depends on a wider clinical and practical context**.

2.4 Measurement modalities and delivery formats

Mental health assessment can be conducted through multiple modalities, each with distinct strengths, limitations, and levels of evidentiary maturity. A taxonomy and evaluation are provided in Appendix C, with summary comparisons in Tables 1, 2, 3, and 4. The strongest and most established evidence base remains with self-report instruments, whether paper or digital, and with clinician-administered interviews, which continue to form the backbone of most current clinical workflows. Computerised cognitive and neuropsychological tasks provide an additional source of objective measurement for selected domains. Other modalities, including gamified assessments, speech and voice analysis, eye tracking, facial and video-based analysis, digital phenotyping, EMA, wearable physiological monitoring, multimodal sensing, neuroimaging, EEG, and NLP-based analysis of clinical or patient-generated text, are active areas of development and may ultimately prove valuable, but in most cases they are not yet sufficiently validated for standalone clinical deployment.

The key point is that no single modality is sufficient across all conditions, populations, and care settings. Digital self-report is scalable, inexpensive, and already aligned with reimbursement and workflow realities, but it remains subjective and is not feasible or appropriate for every patient group (e.g., children or the elderly). Clinician-administered assessment offers greater interpretive depth and diagnostic flexibility, but it is time-intensive, scarce and difficult to scale. Cognitive tasks can add performance-based information, while passive and sensor-based modalities may offer new forms of longitudinal signal, especially for monitoring and risk detection, but they currently face challenges of standardisation, validation, interpretability, privacy, and governance. The choice of modality is also inseparable from deployment context: what is feasible in a specialist clinic or research setting may not be feasible in primary care, on a patient's phone, in a residential facility, or in a correctional environment.

For that reason, the standard should not be built around any single measurement technology. Instead, it should specify **modality-agnostic requirements and define how different modalities can be incorporated according to their evidence tier, risk profile, and intended clinical function.** This supports both present-day deployment using mature methods and future extensibility as newer measurement approaches become clinically validated.

2.5 Care settings and populations

Mental health care is delivered across a wide range of settings, each with its own workflow, constraints, acuity profile, and patient population. A more detailed discussion is provided in Appendix D. These settings include primary care, collaborative care, emergency departments, local mental health centers, acute and palliative care, outpatient psychiatry and specialty mental health services, veterans and military systems, paediatric and adolescent services, residential and long-term care, telehealth and asynchronous care, employee assistance programmes, university and college counselling centres, substance use treatment pathways, forensic and correctional environments, payer and utilisation-management contexts, global and low-resource systems, and consumer-facing wellness and self-managed care platforms. Although these environments differ substantially, many of the same structural problems recur: limited clinician time, fragmented pathways, poor continuity across transitions, inconsistent use of structured assessment, and weak capacity for longitudinal monitoring and measurement-based care.

In addition, capacity constraints are pervasive: in many regions access to specialist care is limited or ab-

sent, with substantial geographic variation in provider availability – in the US alone, several thousand counties are designated Mental Health Professional Shortage Areas – and prolonged wait times for referral-based services, often extending to months or longer (Health Resources and Services Administration, 2024). These constraints mean that even when needs are correctly identified, timely access to appropriate care is not guaranteed, and untreated rates remain high across most disorders even in well-resourced settings (Kohn et al., 2004).

Table 1. Treatment Landscape by Condition — Current Standard of Care, Key Limitations, and Opportunities for Standards-Governed Technology

Condition	Pharmacotherapy (guideline-recommended)	Neuromodulation	Psychotherapy (guideline-recommended)	Key Limitations and Gaps	Opportunity for Standards-Governed Technology
Depression	SSRIs, SNRIs (1st line); TCAs, MAOIs, atypicals, augmentation (2nd+); esketamine, lithium augmentation (TRD) (National Institute for Health and Care Excellence, 2022a)	ECT (TRD, acute suicidality); rTMS (FDA-cleared for TRD); VNS (chronic TRD, FDA-approved); MST (investigational) (National Institute for Health and Care Excellence, 2022a)	CBT, BA, IPT, MBCT (relapse prevention); MBSR/mindfulness; couples therapy; psychodynamic; combined Tx superior to monotherapy (National Institute for Health and Care Excellence, 2022a)	Trial-and-error prescribing; non-response detected late; sexual dysfunction/weight gain undermonitored; publication bias inflates drug effect sizes; therapist shortage limits psychotherapy access (Turner et al., 2008)	MBC to detect non-response early; side-effect monitoring; structured treatment matching; track combined vs mono outcomes
Anxiety disorders	SSRIs, SNRIs (1st line); buspirone (GAD); benzodiazepines (short-term only, dependence risk) (National Institute for Health and Care Excellence, 2011b)	rTMS (emerging evidence for GAD, social anxiety); dTMS (investigational); limited guideline support at present	CBT (exposure-based); ACT; applied relaxation; MBSR/mindfulness; disorder-specific protocols vary (ERP for OCD features, PE for PTSD features) (National Institute for Health and Care Excellence, 2011b)	Benzodiazepine overprescribing; subtype-specific matching rare in practice; long wait for CBT; limited differentiation between anxiety subtypes at point of care (Bandelow et al., 2023)	Multi-domain screening to differentiate subtypes; modality matching; between-session exposure support; benzodiazepine use monitoring
PTSD	SSRIs (sertraline, paroxetine); prazosin (nightmares); limited pharmacological options overall (Schnurr et al., 2024)	rTMS (emerging adjunctive evidence); stellate ganglion block (investigational); limited guideline support	TF-CBT, PE, EMDR (guideline-recommended 1st line); CPT; emerging: MDMA-assisted therapy (regulatory pathway uncertain) (Schnurr et al., 2024)	High dropout from trauma-focused therapy; comorbid SUD/TBI complicate treatment; limited trained therapists; 54% of PTSD trials rated high risk of bias (O'Neil et al., 2024)	Comorbidity-aware treatment matching; dropout risk monitoring; between-session stabilisation; longitudinal symptom tracking
Bipolar disorder	Lithium, valproate, lamotrigine; atypical antipsychotics; complex polypharmacy common (Keramatian et al., 2023)	ECT (acute mania, severe depression, catatonia); rTMS (bipolar depression, emerging); limited evidence for maintenance neuromodulation (Keramatian et al., 2023)	Psychoeducation; CBT adapted; IPSRT; family-focused therapy; adjunctive role only (Keramatian et al., 2023)	Frequently misdiagnosed as unipolar depression; mood monitoring infrequent; lithium monitoring requirements limit use; polypharmacy side-effect burden poorly tracked (Daveney et al., 2019)	Bipolar screening at depression intake; longitudinal mood charting; medication interaction checks; side-effect/lab monitoring alerts
Psychotic-spectrum disorders	Antipsychotics (FGA/SGA); clozapine for treatment resistance; LAIs for adherence (Keepers et al., 2020)	ECT (catatonia, treatment-resistant psychosis); rTMS (auditory hallucinations, emerging evidence); tDCS (investigational) (Keepers et al., 2020)	CBTp; family intervention; social skills training; supported employment; cognitive remediation (National Institute for Health and Care Excellence, 2014)	Metabolic side effects undermonitored; clozapine underused due to monitoring burden; cognitive and social cognitive deficits rarely assessed; substantial life expectancy gap (Correll et al., 2022)	Metabolic monitoring integration; cognitive assessment to inform Tx; adherence tracking; early relapse detection via longitudinal monitoring

Continued on next page

Table 1 continued from previous page

Condition	Pharmacotherapy	Neuromodulation	Psychotherapy	Key Limitations and Gaps	Opportunity for Standards-Governed Technology
OCD	SSRIs (high dose); clomipramine; SGA augmentation (National Institute for Health and Care Excellence, 2005)	DBS (FDA humanitarian device exemption for refractory OCD); rTMS/dTMS (FDA-cleared for OCD); emerging: focused ultrasound (investigational) (National Institute for Health and Care Excellence, 2005)	ERP (gold standard protocol); CBT; ACT (emerging) (National Institute for Health and Care Excellence, 2005)	Average 10+ year treatment delay; ERP therapists scarce; frequently misdiagnosed; comorbid depression complicates treatment sequencing (Dell'Osso et al., 2013)	Earlier detection via routine screening; ERP therapist matching; severity-based stepped care; between-session ERP support
ADHD	Stimulants (methylphenidate, amphetamines); atomoxetine, guanfacine (non-stimulant) (National Institute for Health and Care Excellence, 2018)	FDA-cleared eTNS (paediatric ADHD); rTMS (investigational); neurofeedback (limited guideline support, mixed evidence)	CBT (adult ADHD); coaching; organisational skills training; limited psychotherapy evidence vs pharmacotherapy (National Institute for Health and Care Excellence, 2018)	Adult ADHD under-recognised; long-term outcome data limited; comorbid anxiety/depression frequently untreated; cardiovascular monitoring inconsistent (National Institute for Health and Care Excellence, 2018)	Routine adult ADHD screening; comorbidity detection; treatment response monitoring; cardiovascular safety tracking
Alcohol / substance use disorders	Naltrexone, acamprosate (AUD); buprenorphine, methadone (OUD); varenicline (nicotine); disulfiram (National Institute for Health and Care Excellence, 2011a)	rTMS (emerging evidence for craving reduction in AUD, nicotine dependence); tDCS (investigational); DBS (severe refractory cases, investigational)	MI, CBT, CM, 12-step facilitation; integrated dual-diagnosis treatment rare in practice (National Institute for Health and Care Excellence, 2011a)	MAT grossly under-prescribed (esp. OUD); SUD/MH treated in silos; relapse poorly monitored; abstinence-only bias in much of the literature (Yakovenko et al., 2024)	Integrated MH+SUD screening; MAT decision support; relapse monitoring; harm reduction-compatible care pathways
Eating disorders	Limited pharmacological role; fluoxetine (BN); olanzapine (AN, limited); lisdexamfetamine (BED) (National Institute for Health and Care Excellence, 2017)	rTMS (emerging for AN, BN); tDCS (investigational); DBS (severe refractory AN, investigational)	CBT-E, FBT (adolescents); DBT (BN, BED); MANTRA (AN); nutritional rehabilitation (National Institute for Health and Care Excellence, 2017)	High mortality of any MH condition; severely under-detected in primary care; specialist services scarce; OSFED/BED poorly served by existing trials; PTSD comorbidity high (van Hoeken & Hoek, 2020)	Routine screening (SCOFF) at primary care; comorbidity-aware assessment; specialist referral navigation; outcome monitoring beyond weight
Insomnia	Z-drugs, low-dose trazodone, melatonin agonists; benzodiazepines (not recommended long-term) (Sateia et al., 2017)	Limited evidence; rTMS (investigational for comorbid insomnia); no established neuromodulation protocols for primary insomnia	CBT-I (recommended 1st line in guidelines); digital CBT-I (DTx available) (Edinger et al., 2021)	CBT-I access severely limited despite guideline recommendation; hypnotics prescribed as default; insomnia as transdiagnostic driver under-recognised; sleep rarely monitored longitudinally (Riemann et al., 2017)	Screening insomnia across all MH presentations; digital CBT-I routing; sleep as transdiagnostic treatment target; wearable sleep monitoring integration
Sexual dysfunction	PDE5 inhibitors (EDys); hormonal Tx; antidepressant switching/augmentation for SSRI-induced SD (Parish et al., 2021)	No established neuromodulation protocols for sexual dysfunction	Sex therapy; CBT-based approaches; couples therapy; mindfulness-based interventions (Parish et al., 2021)	Almost never screened in MH settings; iatrogenic SD from SSRIs/antipsychotics under-detected; female SD particularly under-researched; patient reluctance to report without structured prompting (McCabe et al., 2016a)	Routine SD screening alongside MH assessment; medication side-effect monitoring; treatment-emergent SD flagging to prescriber

Continued on next page

Table 1 continued from previous page

Condition	Pharmacotherapy	Neuromodulation	Psychotherapy	Key Limitations and Gaps	Opportunity for Standards-Governed Technology
Autism spectrum	No approved pharmacotherapy for core autism features; medications for comorbidities (irritability, anxiety, ADHD) (National Institute for Health and Care Excellence, 2021)	rTMS (investigational for social cognition and repetitive behaviours); tDCS (investigational); no established clinical protocols	Social skills interventions; adapted CBT for comorbid anxiety/depression; ABA (controversial); occupational therapy; speech-language therapy (National Institute for Health and Care Excellence, 2021)	Late/missed diagnosis (esp. females, adults); standard MH instruments not validated for autistic populations; comorbid MH conditions under-treated; service fragmentation between autism and MH systems (Lai et al., 2019)	Autism-aware screening; adapted assessment pathways; comorbidity detection; coordinated MH care navigation
Suicidality / self-harm	Lithium (bipolar); clozapine (schizophrenia); ketamine (acute ideation, emerging); no broadly approved anti-suicidal medication (National Institute for Health and Care Excellence, 2022b)	ECT (acute suicidality with severe depression); ketamine/esketamine (rapid anti-suicidal effects, emerging); rTMS (adjunctive, limited evidence for suicidality specifically)	DBT; safety planning; CAMS; brief interventions post-attempt; means restriction counselling (National Institute for Health and Care Excellence, 2022b)	Prediction models have low PPV; screening inconsistent across settings; care transitions post-crisis are highest-risk period; treatment of underlying conditions is primary prevention but rarely framed as such (Belsher et al., 2019)	Universal screening at every encounter; safety escalation workflows; care transition monitoring; longitudinal risk tracking across underlying conditions

Abbreviations: ACT = acceptance and commitment therapy; ABA = applied behaviour analysis; AN = anorexia nervosa; AUD = alcohol use disorder; BA = behavioural activation; BED = binge eating disorder; BN = bulimia nervosa; CAMS = Collaborative Assessment and Management of Suicidality; CBT = cognitive behavioural therapy; CBT-E = enhanced CBT for eating disorders; CBT-I = CBT for insomnia; CBTp = CBT for psychosis; CM = contingency management; CPT = cognitive processing therapy; DBS = deep brain stimulation; DBT = dialectical behaviour therapy; dTMS = deep transcranial magnetic stimulation; DTx = digital therapeutic; ECT = electroconvulsive therapy; EDys = erectile dysfunction; EMDR = eye movement desensitisation and reprocessing; ERP = exposure and response prevention; eTNS = external trigeminal nerve stimulation; FBT = family-based treatment; FGA = first-generation antipsychotic; GAD = generalised anxiety disorder; IPT = interpersonal therapy; IPSRT = interpersonal and social rhythm therapy; LAI = long-acting injectable; MANTRA = Maudsley Anorexia Nervosa Treatment for Adults; MAT = medication-assisted treatment; MBC = measurement-based care; MBCT = mindfulness-based cognitive therapy; MBSR = mindfulness-based stress reduction; MH = mental health; MI = motivational interviewing; MST = magnetic seizure therapy; OUD = opioid use disorder; OSFED = other specified feeding or eating disorder; PE = prolonged exposure; PDE5 = phosphodiesterase type 5; PPV = positive predictive value; rTMS = repetitive transcranial magnetic stimulation; SD = sexual dysfunction; SGA = second-generation antipsychotic; SUD = substance use disorder; TBI = traumatic brain injury; tDCS = transcranial direct current stimulation; TF-CBT = trauma-focused CBT; TRD = treatment-resistant depression; Tx = treatment; VNS = vagus nerve stimulation. Guideline-recommended treatments reflect current major guidelines (APA, NICE, WHO, VA/DoD) and do not imply endorsement of the underlying evidence quality.

The differences between settings are also clinically important. In primary care, mental health is usually one of several competing priorities within a brief consultation, which means assessment may need to leverage pre-visit, waiting-room, or portal-based workflows and generate outputs that are actionable within the scope of the primary care provider. This is particularly consequential because primary care physicians prescribe a large share of psychotropic medications despite variable psychiatric training, limited consultation time, and competing clinical priorities. In rural and underserved settings, these pressures are amplified by specialist shortages, longer travel times, and greater reliance on primary care for mental health management, increasing the need for structured diagnostic support that can operate within routine workflows (Katayama et al., 2023; Muench et al., 2022; Waheed et al., 2024).

In emergency departments and acute care, the relevant focus is immediate safety, triage, disposition, and continuity of information at discharge rather than comprehensive multi-domain assessment; the documented gaps in post-discharge follow-up and continuity for psychiatric emergency presentations underline how high the bar is for handoff infrastructure in this setting (Boudreaux et al., 2011). In outpatient psychiatry, specialty care, and telepsychiatry, there is more room for pre-visit digital assessment, severity tracking, treatment response monitoring, and structured decision support, but implementation remains limited by workflow habits and capacity constraints. In EAPs, university counselling centres, and consumer-facing platforms, the challenge is often high-volume, lower-intensity triage and identifying who requires escalation into formal clinical care. In correctional, residential, and long-term care settings, device restrictions, cognitive limitations, proxy reporting, staffing models, and handoff requirements create additional design constraints.

Population-specific factors add a second layer of complexity. Children and adolescents often require multi-informant assessment involving parents, guardians, teachers, or clinicians, since cross-informant agreement is typically low-to-moderate and each informant captures behaviour in a distinct context (De Los Reyes et al., 2015). Neurodivergent individuals may be poorly served by tools designed around neurotypical assumptions and standard self-report formats; for example, alexithymia – difficulty identifying and describing one's own emotions – is highly prevalent in autistic populations and can systematically bias self-report measures of affect (Kinnaird et al., 2019). Older adults may have sensory, motor, or cognitive limitations that affect usability and validity, which is why instruments such as the Geriatric Depression Scale were specifically designed with simplified response formats and reduced reliance on somatic items (Yesavage & Sheikh, 1986). Linguistically and culturally diverse populations raise issues of translation validity, explanatory models of distress, and interpreter use. Veterans and other high-complexity populations often present with overlapping trauma, substance use, traumatic brain injury, and chronic pain – a constellation sometimes referred to as the polytrauma clinical triad and associated with elevated risk of poor outcomes including suicidality (Lew et al., 2009). In many of these cases, standard self-report alone is insufficient, and systems must support clinician-assisted, informant-based, adapted, or otherwise alternative administration routes.

Despite this heterogeneity, **the same core infrastructure needs recur across settings and populations: structured multi-domain assessment, longitudinal monitoring, evidence-informed decision support, and systematic care navigation.** The implication is not that one workflow fits all, but that one underlying standard can be configured across many workflows without being reinvented for each environment. The standard is therefore designed to be setting-agnostic at the level of core requirements, while allowing adaptation at the level of deployment, interface, administration mode, and workflow integration.

2.6 Financial structures, reimbursement, and economic feasibility

Adoption of mental health infrastructure depends not only on clinical validity and technical feasibility but also on economic viability. A more complete analysis is provided in Appendix E. Mental health conditions impose substantial direct and indirect costs through emergency presentations, inpatient admissions, delayed or repeated diagnosis, chronic disability, absenteeism, presenteeism, and wider productivity loss, as well as downstream social costs relating to family functioning, caregiving burden, relationship stability, and child development (Organisation for Economic Co-operation and Development, 2018, 2026; World Health Organization, 2022b). The economic rationale for structured assessment infrastructure is that earlier identification, better treatment matching, more systematic monitoring, and improved care navigation can reduce high-cost escalation while improving outcomes. Conversely, the cost-offset logic is weakened when patients are misclassified, since inappropriate treatment, delayed response detection, and avoidable escalation dilute the expected reductions in physical and psychiatric health care utilisation.

In many jurisdictions, billing codes for behavioural health screening, psychological testing, measurement-based care, chronic care management, and collaborative care models provide a possible route for supporting digital assessment and monitoring workflows within existing financing structures, but fee-for-service reimbursement alone may be unlikely to be sufficient in many settings (Centers for Medicare & Medicaid Services, 2024). Screening codes such as 96127 may be underused because reimbursement is modest and patient cost-sharing

can create implementation concerns (American Medical Association, 2024a; Behave Health, 2024; Prevounce Health, 2026; Rural Health Information Hub, 2025); Collaborative Care Management requires substantial staffing and operational infrastructure and may be cost-neutral or only modestly profitable; and Chronic Care Management reimbursement varies across payers.

By contrast, many emerging modalities such as speech analysis, eye tracking, digital phenotyping, and passive sensing largely remain outside routine reimbursement even where the science is promising. Cost structures also differ markedly across modalities: digital self-report and related measurement-based care workflows are relatively low-cost and scalable; clinician-administered interviews carry the full cost of clinician time; cognitive testing may require specialised platforms or supervision; wearables introduce hardware and data infrastructure costs; and neuroimaging remains far more expensive than any scalable screening modality.

The result is a convergence between clinical practicality, technical scalability, and financial feasibility. The modalities and workflows that are currently most realistic to deploy at scale are also those best aligned with existing reimbursement and cost-offset mechanisms. This does not mean that the standard should exclude more advanced modalities, but it does mean that its **near-term foundation must rest on approaches that can sustain themselves financially in routine care.** Long-term viability depends not only on reimbursement, but also on avoided downstream utilisation, more efficient use of clinician time, support for value-based care and collaborative care workflows, and the secondary value of the structured real-world data generated by compliant systems.

For this reason, the business case for the standard should also be framed through value-based payment models, accountable care arrangements, and quality-reporting incentives. Measurement-based care can generate the outcome data payers need to evaluate quality, support shared-savings models where improved identification and treatment reduce emergency department use or hospitalisation, and create clearer endpoints for psychotherapy and medication management. In this sense, structured assessment infrastructure is not only a billable service layer but also a mechanism for demonstrating value, reducing ineffective utilisation, and aligning behavioural health care with risk-based and outcome-based payment models.

2.7 Strategic prioritisation

The preceding sections describe a landscape that is broad in both the conditions requiring attention and the modalities and treatments available. Not all combinations of condition, measurement approach, treatment category, and care setting are equally ready for implementation, equally impactful, or equally feasible. This section synthesises the dimensions laid out previously to identify where the greatest opportunities and the greatest challenges lie.

The following dimensions are considered in arriving at these priorities:

1. **Prevalence and burden:** Which conditions account for the greatest share of disability, and where does under-detection have the largest downstream consequences?
2. **Treatment availability and decision complexity:** Where do guideline-recommended treatments exist but go under-used, and where is the treatment landscape itself contested, underdeveloped, or complicated by comorbidity?
3. **Measurement maturity and scalability:** Which assessment modalities have established psychometric evidence and can be deployed at population scale, and which remain experimental or setting-bound?
4. **Setting flexibility:** Which approaches are deployable across the widest range of care contexts, and which require specialised infrastructure?
5. **Implementation gap:** Where is the distance between what is already known and what is actually practised the largest—that is, where could standards and technology close a gap that is not primarily a scientific one?
6. **Foundation dependency:** Which capabilities are prerequisites for other capabilities—that is, what must be in place before downstream functions (treatment recommendation, risk stratification, care navigation, AI decision support) can operate?
7. **Evidence generation potential:** Which deployments generate the real-world data needed to address the evidence quality limitations described above?

2.7.1 Where impact is greatest and most achievable

When the dimensions above are applied simultaneously, the relative suitability of different measurement approaches becomes clear. Table 4 evaluates each major modality against the key dimensions.

As Table 4 illustrates, no single modality dominates all dimensions. Digital self-report satisfies the largest number of practical deployment, scalability, and cost criteria, but it is inherently subjective, language- and literacy-dependent, unsuitable for formal diagnostic classification in populations where self-report is unreliable, and provides episodic snapshots rather than continuous measurement. EMA extends self-report into longitudinal and ecologically valid territory and is well-suited for symptom tracking and phenotyping, but is not appropriate for formal diagnostic classification and shares

the subjectivity limitations of all self-report approaches. Clinician-administered interviews offer the greatest diagnostic rigour but are constrained by workforce availability and cost. Cognitive tasks provide objective functional assessment. Wearables offer continuous, passive, objective data with high adherence but lack validated clinical decision thresholds. Speech and eye tracking offer ecologically valid, objective measurement with emerging evidence but face ethical, privacy, cost, and ML-dependency challenges. Neuroimaging provides the highest-resolution data but is facility-bound and not scalable. The implication is that comprehensive mental health assessment requires a layered, multi-modal approach, and that **the standard described in this document must accommodate the full spectrum with evidence-tier requirements calibrated to each modality's maturity and the clinical risk of the function it serves.**

2.7.2 Where the challenges are most complex

At the other end of the spectrum, several condition-treatment-measurement intersections present substantially greater difficulty, and require honesty about what technology-powered systems can and cannot address in the near term.

2.7.3 Treatment-resistant and complex presentations

Treatment-resistant depression, refractory psychosis, severe personality disorders, and complex comorbidity (e.g., PTSD with co-occurring substance use and TBI) involve treatment decisions that guidelines address inadequately, where the evidence base is thinnest, and where clinical judgment is most critical. Decision support in these domains requires sophisticated handling of uncertainty, contraindication cascades, and treatment sequencing logic that cannot be derived from existing trial data alone.

2.7.4 Conditions where self-report is insufficient

Psychotic-spectrum disorders, severe cognitive decline, autism spectrum conditions in those with limited verbal communication, and presentations involving challenges in self assessment (i.e., anosognosia) require assessment modalities beyond self-report - clinician-administered interviews, cognitive tasks, informant reports, and potentially emerging modalities such as speech or gaze analysis. Scalable, validated tools for these populations are limited, and measurement invariance data are sparse.

2.7.5 Conditions where the treatment landscape itself is contested

Autism (where the primary intervention evidence is heterogeneous and politicised), eating disorders (where specialist services may be scarce and the trial base may be dominated by restrictive presentations), and behavioural addictions (where validated screening and treatment are both nascent) represent domains where the evidence needed to inform technology-powered decision support does not yet exist in sufficient quality or breadth.

2.7.6 Emerging measurement modalities

Speech analysis, eye tracking, digital phenotyping, and multimodal sensor fusion hold significant promise but face lack of normative standardisation, and insufficient validation for standalone deployment. The gap here is primarily scientific and engineering-based rather than implementation-based: the technology and methodology need further maturation before standards can meaningfully govern their clinical use.

2.7.7 Populations underserved by existing instruments

Children and adolescents, older adults with cognitive impairment, non-English-speaking populations, justice-involved individuals, and people with intellectual disabilities are poorly served by instruments validated primarily on English-speaking adult communities or clinical samples. Measurement invariance data for these groups are limited, and cutoff thresholds may not generalise. Further, lack of cultural competency on the part of interviewers can be a major barrier. Standards must acknowledge these gaps explicitly rather than assume universal applicability of existing tools.

Table 2. Measurement Modalities — Evidence Maturity, Scalability, and Primary Settings

Modality	**Evidence Maturity**	**Scalability**	**Primary Feasible Settings**
Paper-and-pencil self-report (Kroenke et al., 2001)	Established	High	All settings; especially low-resource, no technology required, but high-risk and high-effort for digitalization.
Computer/tablet questionnaires (Gibbons et al., 2008)	Established	High	Primary care, psychiatry, telehealth, EAPs, university clinics, research
Mobile (smartphone) self-report (Torous & Roberts, 2017)	Established	Very high	Remote/asynchronous, between-session, community, pre-visit
Clinician-administered interviews (First et al., 2016)	Established (gold standard)	Low	Psychiatry, specialty MH, forensic, diagnostic confirmation
Computerised cognitive tasks (Gibbons et al., 2008)	Established (research); growing (clinical)	Moderate	Psychiatry, neuropsychology, research, controlled digital
Gamified / serious game assessments (Lumsden et al., 2016)	Emerging	High (potential)	Paediatric, ADHD, schools, remote, engagement-sensitive
Speech and voice analysis (Low et al., 2020)	Emerging	High (potential)	Telehealth, phone-based, clinical encounters, remote
Eye tracking and gaze analysis (Armstrong & Olatunji, 2012)	Emerging	Low–moderate	Lab/research, specialty clinic; emerging via tablets/webcams
Facial expression / video analysis (Girard & Cohn, 2015)	Emerging	Moderate (potential)	Telehealth (video), lab/research, clinical encounters
Multimodal sensor fusion (Mohr et al., 2017)	Experimental	Low	Lab/research, controlled clinical, specialised platforms
Digital phenotyping (passive sensing) (Insel, 2017)	Emerging	Very high (potential)	Continuous remote, between-session, longitudinal monitoring
EMA / experience sampling (Shiffman et al., 2008; Trull & Ebner-Priemer, 2013)	Established (research); growing (clinical)	High	Mobile/wearable, between-session, daily life, research
Wearable physiological monitoring (Hickey et al., 2021)	Emerging	High	Remote, residential, continuous, daily life
Continuous inpatient monitoring (Mohr et al., 2017)	Emerging	Low (setting-bound)	Inpatient, residential, intensive outpatient
Neuroimaging (MRI, fMRI, PET) (Etkin, 2019)	Established (research); limited (clinical MH)	Very low	Academic medical centres, specialist neurology/psychiatry
EEG / neurophysiological assessment (Newson & Thiagarajan, 2019)	Established (research); emerging (clinical MH)	Low	Neurology, specialty psychiatry, research; emerging consumer-grade
NLP + LLMs for clinical / patient text (Le Glaz et al., 2021)	Emerging	High (potential)	EHR integration, telehealth, chatbot, journaling, therapy transcripts

Citations

indicate representative reviews or seminal references for each modality. Evidence maturity: Established = multiple systematic reviews, validated for clinical use; Emerging = growing primary literature, not yet validated for standalone clinical deployment; Experimental = proof-of-concept, limited clinical validation. Scalability: reflects feasibility of population-level deployment given current technology and cost constraints.

Table 3. Condition Domains × Applicable Measurement Modalities

Condition Domain	**Primary Modalities (Established)**	**Emerging / Adjunctive Modalities**
Depression	Self-report (PHQ-9), clinician interview, EMA	Speech, digital phenotyping, wearable (sleep/activity), NLP, cognitive tasks, EEG, fMRI
Anxiety disorders	Self-report (GAD-7, Mini-SPIN, PDSS-SR), clinician interview, EMA	Wearable (HRV, EDA), digital phenotyping, speech, eye tracking
PTSD	Self-report (PCL-5), clinician interview	EMA, wearable (sleep, HRV), speech, eye tracking (attentional bias), digital phenotyping, fMRI
ADHD	Self-report (ASRS v1.1), clinician interview, cognitive tasks	Gamified assessment, eye tracking, digital phenotyping, EEG, actigraphy
Bipolar spectrum	Self-report (ASRM), clinician interview, EMA	Digital phenotyping (activity/sleep cycles), speech, wearable, actigraphy
Psychotic-spectrum	Self-report (CAPE), clinician interview, cognitive tasks (WCST), emotion recognition	Speech, digital phenotyping, NLP, fMRI, EEG
OCD	Self-report (Y-BOCS), clinician interview	EMA, cognitive tasks (set-shifting), digital phenotyping (behavioural patterns)
Alcohol / substance use	Self-report (AUDIT, DAST), clinician interview	EMA, digital phenotyping (mobility/social), wearable (sleep), cognitive tasks, NLP
Eating disorders	Self-report (SCOFF), clinician interview	EMA, cognitive tasks (set-shifting), NLP (journal analysis)
Insomnia / sleep	Self-report (ISI), clinician interview	Wearable actigraphy, digital phenotyping (screen use, circadian), EMA, EEG (polysomnography)
Sexual dysfunction	Self-report (FSFI, IIEF-5, ASEX), clinician interview	EMA (real-time symptom tracking)
Autism spectrum	Self-report (AQ-10), clinician interview, cognitive tasks	Eye tracking (social attention), emotion recognition, gamified, speech prosody, EEG, fMRI
Cognitive decline	Self-report (AD8), clinician interview, cognitive tasks (TMT)	Gamified cognitive, speech, eye tracking, digital phenotyping (mobility), MRI, EEG
Suicidality / self-harm	Self-report (ASQ, C-SSRS), clinician interview	NLP (clinical notes, text), EMA, digital phenotyping, speech, continuous monitoring

Primary modalities are those with established clinical evidence sufficient for deployment. Emerging/adjunctive modalities have research support but have not yet reached the evidence threshold for standalone clinical use; they are positioned as complementary to primary modalities. Specific instruments referenced are illustrative, not exhaustive.

Table 4. Measurement Modality Suitability Matrix — Cross-Cutting Assessment Dimensions

Dimension	Digital self-report	EMA	Clinician interview	Cognitive tasks	Wearables	Speech / voice	Eye tracking	Neuro-imaging
Evidence and validity								
Established evidence for symptom detection or diagnostic classification	✓	∼	✓	✓	×	×	×	∼
Suitable for formal diagnostic screening or classification	✓	∼	✓	∼	×	×	×	∼
Suitable for longitudinal symptom tracking and phenotyping	✓	✓	∼	✓	✓	∼	∼	×
Covers high-prevalence conditions (depression, anxiety, PTSD, SUD, insomnia)	✓	✓	✓	∼	∼	∼	∼	∼
Objective measurement (free from self-report bias, social desirability, recall distortion)	×	×	×	✓	✓	✓	✓	✓
High signal-to-noise ratio per individual administration	✓	×	✓	✓	×	∼	∼	✓
Interpretable without pre-trained ML models or large reference datasets	✓	∼	✓	✓	×	×	×	∼
Usable when self-report is unreliable (psychosis, anosognosia, cognitive impairment, young children)	×	×	✓	✓	∼	∼	✓	✓
Language- and interpretation-independent	×	×	×	✓	✓	∼	✓	✓
Does not require literacy or digital literacy	×	×	✓	∼	✓	✓	✓	✓
Captures cognitive and neuropsychological function	×	×	∼	✓	∼	✓	✓	✓
Captures naturalistic, real-world functioning (ecological validity)	×	✓	×	×	✓	✓	✓	×
Supports digital phenotyping (contributes to continuous behavioural or physiological characterisation)	∼	✓	∼	✓	✓	✓	✓	✓
Temporal resolution and coverage								
Supports longitudinal measurement-based care	✓	✓	∼	✓	✓	∼	∼	×
Continuous or near-continuous monitoring	×	∼	×	×	✓	×	×	×
Administrable asynchronously (outside of a clinical encounter)	✓	✓	×	∼	✓	✓	✓	×
Potential for early or pre-symptomatic detection	∼	∼	∼	∼	∼	∼	∼	∼
Practical feasibility and scalability								
No specialised hardware required	✓	✓	✓	✓	∼	✓	✓	×

Continued on next page

Table 4 continued from previous page

Dimension	Digital self-report	EMA	Clinician interview	Cognitive tasks	Wearables	Speech / voice	Eye tracking	Neuro-imaging
No clinician time required for administration	✓	✓	×	✓	✓	✓	✓	×
No controlled environment required	✓	✓	∼	✓	✓	✓	✓	×
Low patient burden per administration	✓	✓	×	∼	✓	∼	∼	×
High adherence and completion rates in practice	∼	∼	∼	∼	✓	∼	∼	×
Setting feasibility								
Feasible: primary care, EAP, university, workplace	✓	∼	×	×	∼	∼	∼	×
Feasible: emergency and acute care	✓	×	∼	×	×	×	×	×
Feasible: telehealth and remote care	✓	✓	✓	∼	✓	∼	∼	×
Feasible: low-resource and global settings	✓	∼	×	×	×	✓	×	×
Feasible: correctional, forensic, justice-involved	✓	×	∼	×	×	×	×	×
System-level value								
Provides structured data for downstream CDS, risk stratification, care navigation	✓	∼	✓	∼	×	×	×	×
Generates real-world data for evidence revalidation at scale	✓	✓	∼	∼	✓	∼	∼	×
Ethical, legal, and cost considerations								
Minimal ethical/privacy concerns for routine deployment	✓	∼	∼	∼	∼	×	×	×
Favourable licensing and IP terms available	∼	∼	∼	∼	✓	✓	✓	∼
Low implementation and per-administration cost	✓	✓	×	∼	∼	∼	∼	×
Existing reimbursement or payer coverage pathways	✓	×	✓	∼	×	×	×	✓
Evidence for reduction in downstream care costs	✓	∼	✓	∼	∼	×	×	∼

✓ = criterion met; ∼ = partially met or context-dependent; × = criterion not met. MBC = measurement-based care; CDS = clinical decision support; EMA = ecological momentary assessment. Digital phenotyping is included as a dimension ("supports digital phenotyping") rather than a column because it represents a data-processing and characterisation paradigm applied to data from multiple assessment modalities rather than a standalone assessment method. No single modality dominates all dimensions.

2.7.8 Personalised treatment sequencing

The long-term vision of dynamically optimised, individual-level treatment selection - choosing not just a treatment category but a specific agent, dose, modality, and sequencing strategy for a specific patient - requires causal inference methodology (dynamic treatment regimes, g-computation) that the current evidence base cannot support at scale. This represents a genuinely hard scientific problem, not an implementation gap, and standards should be clear-eyed about the distance between current capability and this aspiration. One can typically establish whether a treatment works on average, or attempt to identify for whom it works, but not both with high reliability in the same study. Designs that are optimal for one objective are usually poorly suited to the other, and studies capable of doing both well are rare and difficult to conduct.

The distinction between these two categories - implementation gaps that standards and technology can close now, and scientific gaps that require further research before standards can meaningfully govern them - is central to the framework proposed in this document. The evaluation tier system is designed to make this distinction operational: different levels of clinical claim require different levels of evidence, and the standard explicitly accommodates the full spectrum from well-validated screening to experimental modalities, each at an appropriate evidence threshold.

3 The proposed technology-powered reference standard

The preceding section established the landscape: the conditions, treatments, measurement modalities, care settings, financial structures, and prioritisation logic that any standard must address. This section defines the standard itself. Before specifying its components, we state the design constraints that any viable proposal must satisfy simultaneously, then present the reference model, evaluation framework, and governance structure.

3.1 Design constraints and feasibility requirements

For the proposed standard to achieve adoption, it must pass a feasibility test defined by five simultaneous constraints. **First, seamless workflow integration:** the infrastructure must fit within existing clinical workflows through deep EHR integration, not require separate tools or logins. In primary care, for example, this means assessment results embedded in the encounter note before the clinician enters the room. In the ED, it means triage-compatible rapid screening. In telehealth, it means pre-visit digital assessment delivered asynchronously.

Second, reimbursability: the infrastructure must not cost clinics money to operate. As described in Section 2.6, reimbursement pathways for structured screening and measurement-based care already exist (Current Procedural Terminology code 96127, CoCM codes 99492–99494, chronic care management codes), and the proposed standard is designed around activities that are already billable.[1]

Third, low cost and high scalability: this rules out modalities requiring expensive hardware, specialist facilities, or extensive clinician time as the primary assessment layer.

Fourth, applicability across the widest range of populations and settings: while no single approach serves every population equally well, the infrastructure must cover the broadest possible base and define adapted pathways where the primary approach falls short - paediatric populations with neurodevelopmental disorders, older adults with cognitive impairment, non-English-speaking populations, and others.

Fifth, honest acknowledgement of limitations: no single infrastructure can solve every mental health care problem simultaneously. The standard prioritises the highest-impact, most feasible starting point while explicitly defining what it does not address.

Overall, adoption barriers are not solely technical or financial; clinicians may overestimate the reliability of unstructured assessment and may perceive structured measurement as reductive, even where evidence demonstrates improved consistency and outcomes. A related and widespread bias is the assumption that patients will not disclose sensitive information on standardised instruments - particularly regarding substance use, psychotic symptoms, suicidal ideation, and self-harm behaviours. While under-reporting is a real phenomenon, it is not unique to structured assessment; disclosure rates on validated instruments are generally comparable to or higher than those obtained through unstructured clinical inquiry, particularly when instruments are administered in private, low-pressure formats such as pre-visit digital questionnaires. This is important because screening implementation itself varies substantially across health systems and is often limited to brief in-person encounters or metric-driven workflows. Emerging evidence suggests that screening performed primarily to satisfy operational requirements may still fail to identify a substantial proportion of patients with clinically significant symptoms, particularly when patients are uncomfortable disclosing sensitive experiences in hurried or low-trust interactions. Systematic deployment may also improve consistency and equity of detection

[1] We acknowledge that reimbursement in practice is heterogeneous and operationally complex, and payment levels for key components are modest (e.g. Medicare reimbursement for CPT 96127 is typically approximately $4–$7 per instrument, with a national average of around $5 per unit depending on locality and year; Centers for Medicare & Medicaid Services, 2024), so implementations must be explicitly designed to support viable care models rather than assuming reimbursement is automatic.

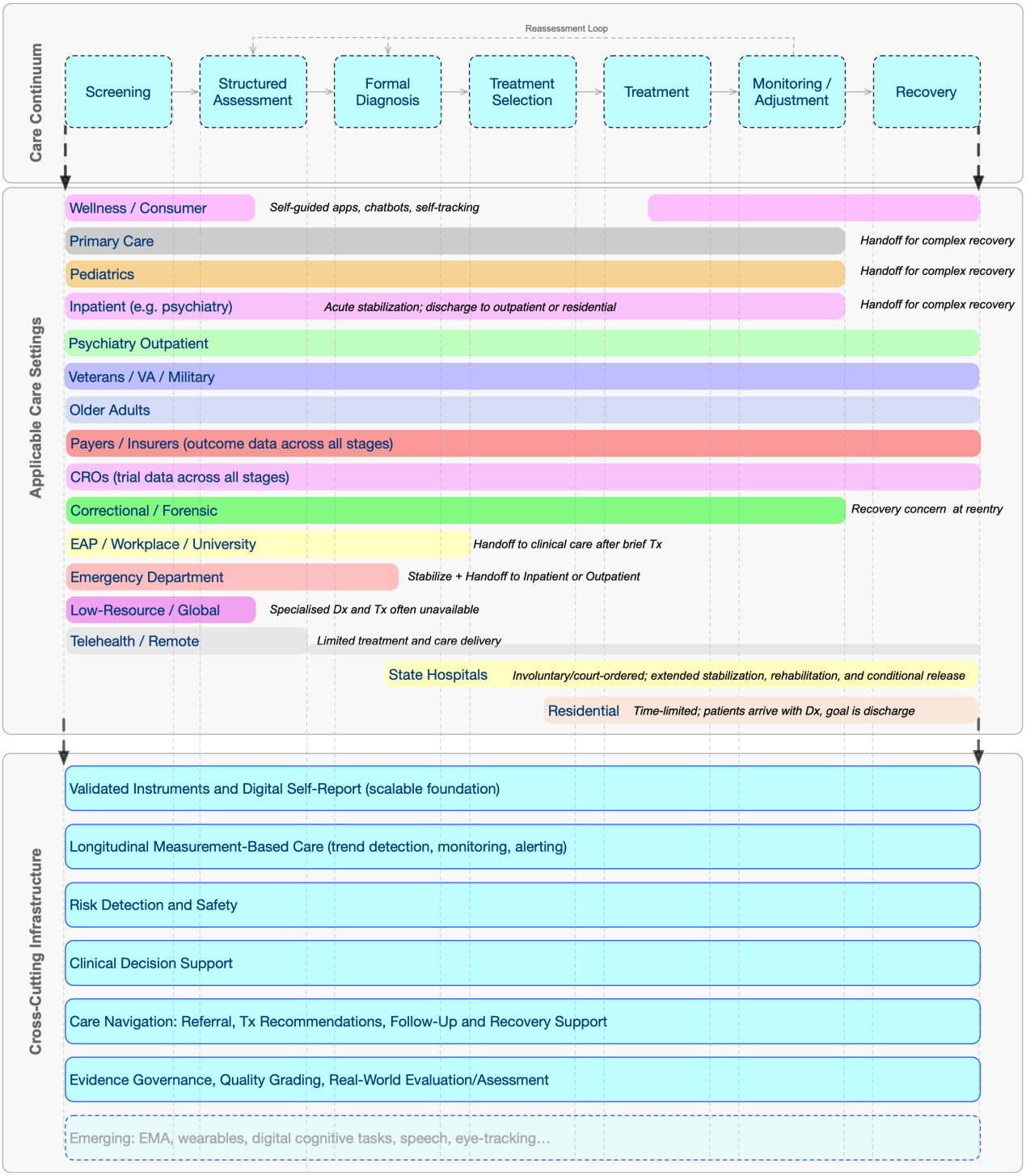

Figure 1. Illustrating how the reference standard cuts across the care continuum and across the different care settings.

relative to selective or clinician-dependent screening practices, while increasing overall screening completion rates (Garcia et al., 2022).

In primary care, for instance, the use of validated instruments is less frequent, so **increased utilisation of validated scales has the potential to significantly improve diagnostic accuracy**, notwithstanding the concomitant need to integrate this administration into the workflow without increasing clinician burden. Finally, it is worth noting that this standard explicitly separates structured measurement from clinical decision authority: systems are responsible for generating reproducible, evidence-linked assessments, while diagnostic and treatment decisions remain the responsibility of the clinician. **Structured measurement augments but does not replace the clinical interaction; rapport, contextual judgment, and the capacity to explore ambiguity and nuance in real time remain essential elements of competent care that no instrument can fully replicate.**

3.2 The reference model

The reference model defines several integrated capabilities that systems in this class should support. Conceptually the approach is depicted in Figure 1. These are not features of a particular product; they are the components of a reference architecture against which any implementation can be evaluated.

Structured, multi-domain assessment. Deployment of validated screening instruments across diagnostic domains using adaptive routing and branching to minimise respondent burden while maintaining sensitivity. The instrument administration should incorporate brief instruments (4-20 items), with adaptive branching that reduces the typical respondent experience to a fraction of a full-battery approach and restricts the assessment time to less than 5 minutes on average.

Longitudinal measurement-based care. Scheduled reassessment, trend detection, and clinician alerting that transforms screening from a one-off event into a continuous monitoring process. Instruments with demonstrated responsiveness to clinical change enable earlier detection of treatment non-response and timely intervention adjustment. The assessment cadence must be configurable per setting and per patient acuity.

Risk detection and safety escalation. Real-time stratification across suicide, self-harm, violence, substance use, and functional decline, with clinician notification for high-risk outputs and recommended escalation pathways. All clinical outputs are advisory; the clinician retains diagnostic and treatment authority at all times.

Evidence-based clinical decision support. Treatment recommendations (pharmacological, adjunct and non-pharmacological) incorporating clinical guidelines, patient profile, contraindications, prior treatment history, and patient preferences, grounded in an evidence library with strength-of-evidence grading and provenance tracking.

Care navigation and continuity. Bridging the transition between assessment completion and treatment initiation through referral recommendations, potential for wait-time support (e.g. via a chatbot) and between-session exercises, and longitudinal follow-up through care transitions and into sustained recovery.

Interoperability and auditability. Deep integration with EHR systems is a non-negotiable requirement for adoption. Systems must some level of bidirectional integration support via IT interoperability standards at the data-level (structured assessment data exchangeable via FHIR resources), workflow-level, and presentation-level (embedded UI components that render within the EHR). Every clinical output must be logged with full provenance: input data, algorithm version, evidence sources, timestamp, and clinician review status. Audit logs must be immutable and retained for the regulatory minimum.

Evidence governance and continuous validation. A governed evidence layer that treats validated instruments and clinical guidelines as the best available starting point, deployed provisionally with active performance monitoring. This includes quality grading, provenance tracking from primary study through clinical guideline to system output, and continuous revalidation using real-world deployment data - prospective diagnostic accuracy estimation, instrument comparison, subgroup fairness analysis, and treatment outcome tracking.

3.3 Safety, privacy, and equity requirements

Mental health data is among the most sensitive categories of health information, and systems operating in clinical mental health contexts can cause direct harm through errors. The standard defines proportionate safety, privacy, and equity requirements.

Three levels of human oversight are defined: autonomous operation with audit trail (low-risk informational functions), clinician notification with override capability (moderate-risk screening), and mandatory clinician review before action (high-risk safety and prescribing functions). No system may autonomously initiate treatment, modify medication, or discharge a patient.

Systems must implement encryption at rest and in transit, role-based access controls, and multi-factor authentication. Informed consent must be granular, comprehensible, and revocable, with separate consent for clinical use, research, and raw data retention. Mental health data protection requirements vary across jurisdictions, and systems must be configurable accordingly.

Systems must support assessment in multiple languages using psychometrically validated translations (not simply machine translation). Instruments must be available at appropriate reading levels with audio-assisted delivery options. Interfaces must comply with accessibility standards. Offline and low-bandwidth modes are required for settings with limited connectivity. For populations where digital self-report is inapplicable - young children, individuals with severe cognitive impairment, those without device access - the standard defines alternative administration pathways and paper-equivalent workflows that produce the same structured data outputs.

3.4 Regulatory positioning, implementation, and governance

The regulatory classification of technology-powered mental health tools depends on the clinical claims they make, not on the technology they use. This distinction is critical because it defines a substantial opportunity space for innovation that does not require the time, cost, and risk of SaMD regulatory clearance. Under the 21st

Century Cures Act (US) and equivalent frameworks, clinical decision support tools that meet four criteria - they do not acquire signals from a nonexempt medical device, they are intended to display or analyse clinical information, they enable independent clinician review, and the clinician is not primarily relying on the software - are exempt from device regulation provided that transparency, interpretability, and clinician independence are demonstrably maintained.

The core functions of the proposed standard (structured screening, severity scoring, treatment option presentation, guideline-based recommendations) may be designed to operate within this exemption. This means that health systems, developers, and innovators can deploy clinically meaningful infrastructure without (at least initially) navigating FDA pre-market review, provided they maintain the transparency and clinician-independence requirements. For higher-risk functions (autonomous diagnostic classification, treatment selection algorithms that go beyond presenting options), SaMD classification applies, and the standard maps these functions to the appropriate regulatory pathway.

Overall, implementation requires mapping the infrastructure onto existing clinical workflows rather than asking clinicians to change their practice. The standard defines workflow integration points for each care setting: pre-visit digital assessment (primary care, psychiatry, telehealth), triage-integrated screening (ED), intake assessment (EAP, university, correctional), between-session monitoring (all settings), and discharge/transition assessment.

3.5 Alternative approaches and counterpositions

The convergence analysis presented above leads to a broad architectural conclusion. However, several alternative perspectives warrant consideration.

One position holds that experienced clinicians (particularly those in primary care) can rely on unstructured assessment without the need for standardised instruments, and that formal measurement risks oversimplifying complex clinical presentations. Clinical expertise is clearly indispensable, particularly for differential diagnosis and high-complexity cases. However, unstructured assessments are inherently variable, difficult to reproduce, and not scalable across settings characterised by limited time and heterogeneous training. This variability is compounded by substantial differences in mental health training across disciplines and specialisations: credentialed mental health clinicians typically receive more extensive preparation than primary care physicians, and within primary care itself, the depth of behavioural health training varies considerably - for example, between family medicine and internal medicine residency programmes. The limitations outlined above - delayed detection, misclassification, and inconsistent treatment pathways - are in part structural consequences of this variability. Standardised measurement does not replace clinical judgment, but provides a reproducible foundation on which it can operate.

A further advantage is that structured assessment infrastructure allows specialist clinicians to focus on the aspects of care that most depend on advanced clinical expertise: differential diagnosis, management of treatment-resistant or high-risk presentations, complex comorbidity, psychotherapy, and nuanced treatment planning. Routine screening, symptom tracking, and longitudinal measurement can be systematised without displacing the uniquely interpretive and relational components of psychiatric and psychological care.

A second perspective emphasises the role of emerging objective modalities, including digital phenotyping, speech and voice analysis, neuroimaging, and multimodal sensor fusion, as potential replacements for self-report. These approaches are promising and may play an important role in future systems, particularly for populations in which self-report is unreliable. At present, however, most remain limited by incomplete validation, lack of standardisation, dependence on machine learning models trained on restricted datasets, and absence of established clinical decision thresholds. Their current role is therefore more appropriately conceptualised as adjunctive to, rather than substitutive for, validated assessment approaches.

A related argument is that condition-specific tools are sufficient for care delivery, with different systems addressing different disorders. In practice, mental health conditions are highly comorbid and frequently present with symptoms that do not fully meet criteria for discrete diagnoses. Fragmented, condition-specific tools risk reproducing the siloed structure of existing care systems and do not support integrated assessment, treatment selection, or care navigation across domains.

Recent advances in large language models have also led to the suggestion that conversational systems could replace structured assessment by eliciting clinically relevant information in natural language. While such systems may capture rich qualitative data, they currently lack the standardisation and psychometric grounding required for reliable measurement, benchmarking, and longitudinal comparison. Trained models have been used to rescore structured interviews, such as the HAM-A and MADRS. However, these interviews have highly structured questions and anchors, without structured anchors, scoring strategies are very challenging to validly implement. Conversational approaches may augment assessment, but do not currently provide a substitute for validated measurement frameworks.

Finally, it may be argued that improvements in clinician training, workforce expansion, and care delivery models could address many of the limitations identified without the need for new technological infrastructure. While these efforts are necessary, obtaining funding for

these changes seem quite unlikely in the foreseeable future and the constraints described - time-limited encounters, variability in assessment practices, and lack of standardized data - are structural features of the system that are unlikely to be resolved through training alone. Measurement infrastructure addresses these constraints directly and operates as a complement to, rather than a replacement for, clinical expertise and workforce development.

Taken together, these perspectives highlight important directions for future research and system development. **However, when evaluated against the combined constraints of evidence maturity, scalability, cost, workflow integration, and regulatory feasibility, these alternatives do not currently provide a viable foundation for population-level deployment.** The reference model proposed in this paper is therefore positioned not as the only possible architecture, but as the most defensible starting point under present conditions.

4 Conclusion and call to action

This paper reviewed the challenges associated with diagnostic assessments, selection of appropriate treatments, and monitoring of their efficacy across conditions, treatments, measurement modalities, care settings, and financial structures. Section 2 documented six structural challenges: low detection rates, high misdiagnosis rates, episodic rather than longitudinal assessment, inconsistent care navigation, technology outpacing governance, and limitations in the evidence base itself. It surveyed the proposed clinical scope (21 condition and presentation domains), the treatment modalities (five categories with their respective evidence limitations), the available validated measurement approaches (eight modality families evaluated across 30+ dimensions), and the care settings in which these conditions present (17 settings from primary care to consumer wellness, each with distinct workflows and constraints).

The clinical case for structured screening and measurement-based care rests on improving detection, reducing misdiagnosis, and enabling timely treatment adjustments — outcomes that translate directly into better patient experiences, better clinical outcomes, and fewer escalations to high-acuity settings. The financial analysis established that reimbursement pathways for these approaches already exist and that cost structures favour digital self-report over strategies such as digital phenotyping and chatbot assessment whose evidence base has not yet matured sufficiently for population-level deployment, reinforcing rather than driving the clinical rationale.

The convergence analysis evaluated measurement modalities against multiple dimensions simultaneously - evidence maturity, cost, scalability, reimbursability, setting flexibility, and workflow integration - and **identified validated digital self-report as one of the only currently deployable approaches that satisfies these constraints at population level,** while documenting its limitations: subjectivity, language dependence, and inapplicability to populations where self-report is unreliable. Section 3 defined a reference standard built on this foundation: a vendor-neutral reference model specifying seven integrated capabilities, an evaluation framework tiered by clinical claim and risk, safety and privacy requirements proportionate to data sensitivity, equity requirements addressing the populations the primary approach does not serve, and a regulatory positioning that identifies the CDS exemption as the primary entry point for innovation. The standard defines a layered approach in which digital self-report serves as the scalable foundation, complemented by clinician-administered assessment, cognitive tasks, and emerging objective modalities, each governed by evidence-tier requirements proportionate to their clinical risk.

The financial analysis identified existing reimbursement pathways (Current Procedural Terminology screening codes, CoCM codes), cost-offset mechanisms through reduced downstream requirement on human based structured assessment , and real-world evidence generation as the three pillars of financial sustainability. The regulatory analysis identified a substantial space within the CDS exemption where clinically meaningful screening and decision support tools can be deployed without SaMD classification. The safety, privacy, and equity sections defined the constraints within which compliant systems must operate, and the governance framework defined how the standard itself will be maintained and updated.

The standard presented here is a preliminary proposal. It is intended to be tested against real-world implementation, challenged on its assumptions and priorities, and revised through the governed versioning and community input process. The multidisciplinary authorship reflects the range of expertise required to develop, validate, and govern this infrastructure.

Call to action: The implementation of a reference standard for technology-powered mental health assessment depends on coordinated roles across the ecosystem. Professional bodies and clinical organisations are responsible for defining practice norms and ensuring that evaluation frameworks are reflected in guidelines and standards of care. Researchers and academic institutions play a central role in addressing the identified evidence gaps, including diagnostic accuracy, measurement invariance, comparative effectiveness, and real-world validation. Developers and vendors are responsible for transparent reporting of clinical claims, evidence sources, validation populations, and system limitations, as well as for maintaining safety, fairness, and auditability in deployed systems. Health systems and payers are responsible for adopting evaluation cri-

teria in procurement and reimbursement decisions, and for supporting the generation of real-world evidence through routine data collection and monitoring. Regulators are responsible for clarifying classification boundaries, particularly for clinical decision support, and for ensuring that oversight mechanisms are proportionate to risk. Finally, patients and caregivers contribute to the governance and evolution of these systems through engagement, feedback, and lived experience, ensuring that implementation remains aligned with real-world needs.

About the authors and disclosures

The authors declare that this manuscript reflects their independent academic and clinical perspectives. The views expressed do not necessarily represent the official positions of their affiliated institutions.

Dr. Matthew Vowels, PhD Eng., PhD Appl. Math.

Chief Technology Officer, Kivira Health; Senior Researcher, The Sense (CHUV & HES-SO Valais); Research Fellow, University of Surrey and University of Roehampton

Matthew Vowels is Chief Technology Officer at Kivira Health, where he leads the development of technology-powered mental health systems at the intersection of clinical care, machine learning, and real-world deployment. He holds two PhDs, in Applied Mathematics from the University of Lausanne and in Engineering from the University of Surrey.

His work focuses on developing and translating advances in AI, causal inference, and statistical modeling into clinically usable infrastructure, spanning multimodal assessment, evidence quality evaluation, conversational systems, and scalable platforms for sensitive health contexts.

Disclosures: As well as working for Kivira Health, he is also an options holder.

Prof. Charles B. Nemeroff, MD, PhD

Chair, Department of Psychiatry and Behavioral Sciences, Dell Medical School, University of Texas at Austin

Charles B. Nemeroff is a leading figure in academic and biological psychiatry. He serves as Chair of Psychiatry and Behavioral Sciences at Dell Medical School, where he also directs the Institute for Early Life Adversity Research and co-directs the Center for Psychedelic Research and Therapy.

A member of the National Academy of Medicine, his work has been central to advancing understanding of mood disorders, trauma, and psychopharmacology.

Disclousures: He is supported by the National Institutes of Health, the National Institute of Mental Health, the Texas Child Mental Health Consortium and the National Institute of Alcohol Abuse and Alcoholism. He is a consultant for Abbott Laboratories, Engrail Therapeutics, Sero (previously Galen Mental Health LLC), Senseye Inc, Precisement Health, Autobahn Therapeutics Inc, EMA Wellness, Denovo Biopharma LLC, Alvogen, Acadia Pharmaceuticals, Inc, Reunion Neuroscience, Kivira Health, Inc, Wave Neuroscience, Invisalert Solutions Inc., Bluecarta Inc., Incannex Healthcare Inc., Supernus, Concierge Clinical Trials Inc., Otsuka, Sensorium, Sirtsei Pharmaceuticals Inc, and Neurocrine Biosciences, LLC.

Charles Nemeroff owns the following patents: Method and devices for transdermal delivery of lithium (US 6,375,990B1), Method of assessing antidepressant drug therapy via transport inhibition of monoamine neurotransmitters by ex vivo assay (US 7,148,027B2), Compounds, Compositions, Methods of Synthesis, and Methods of Treatment (CRF Receptor Binding Ligand) (US 8,551, 996 B2). Charles Nemeroff owns stock in Corcept Therapeutics Company, EMA Wellness, Precisement Health, Relmada Therapeutics Inc, Signant Health, Galen Mental Health LLC, Kivira Health, Inc., Denovo Biopharma LLC, Headlamp Health Inc. Bluecarta Inc, and Senseye Inc.

Dr. Ish Bhalla, MD

Yale University; Duke University, USA

Ish Bhalla is a physician focused on behavioural health integration, complex care delivery, and value-based healthcare systems. His work spans primary care, population health, and care models for medically and socially complex populations, with particular emphasis on integrating mental health assessment and longitudinal care within real-world clinical workflows.

He has experience across health system operations, multidisciplinary care coordination, and technology-enabled service delivery, with interests in scalable behavioural health infrastructure, access to care, and implementation within resource-constrained settings.

Disclosures: Dr. Ish Bhalla is an options holder in Kivira Health.

Disclosures: Dr. Bhalla is employed by WelbeHealth and serves as an advisor to SimplePractice.

Prof. Philip D. Harvey, PhD

Leonard M. Miller Professor of Psychiatry; Director, Division of Psychology, University of Miami Miller School of Medicine

Philip D. Harvey is a leading expert in cognition, functioning, and outcomes measurement, with a focus on digital strategies, in serious mental illness. His work has shaped how psychiatric conditions are evaluated and treated beyond a symptom-level focus, particularly in relation to real-world functioning and disability.

Disclosures: Dr. Harvey has received consulting fees or travel reimbursements from Kivira Health, Alkermes, BMS (Karuna Therapeutics), Boehringer Ingelheim, Kynexis, Minerva Neurosciences, Neurocrine Biosciences, and Recognify. He has a research grant from Intracellular Therapeutics (Now part of Johnson and Johnson). He receives royalties from the Brief Assessment of Cognition in Schizophrenia (Owned by Clario, Formerly WCG Endpoint Solutions, Inc, Formerly Verasci, and contained in the MCCB). He is chief scientific officer of i-Function, Inc and Scientific Consultant to EMA Wellness, Inc.

Dr. Hank Capps, MD, FAAFP

Founding Partner, Seny Advisory

Hank Capps is a physician-executive with extensive experience in technology and healthcare transformation. He has led large-scale clinical, operational, and digital initiatives within major provider organisations, with a focus on implementation and system transformation. He most recently served as Executive Vice President, Chief Information & Digital Officer and President, Catalyst by Wellstar at Wellstar Health System.

Disclosures: Dr. Capps is an options holder in Kivira Health.

Prof. Carol L. Alter, MD

Executive Associate Chair and Professor, Department of Psychiatry and Behavioral Sciences, Dell Medical School, University of Texas at Austin

Carol L. Alter, MD, is the Executive Associate Chair, and Professor of Psychiatry at the University of Texas-Austin, Dell Medical School. She has over 30 years experience as a consult-liaison psychiatrist and expertise in integration of behavioral and physical healthcare and value based care delivery She has been engaged in research regarding these topics, has helped to implement a number of Collaborative Care programs across both public and private settings and has been an advocate for policy reforms aimed at insuring reimbursement and access to integrated and collaborative care.

Prior to joining Dell Medical School, she served as the System Lead for Behavioral Health at Baylor Scott and White Health responsible for the development of behavioral health services, including the Collaborative Care Model (CoCM) delivered to over 2 million primary care patients. She also served as the Chief Medical Officer of Mindoula Health, the first national provider of technology enabled CoCM. She led efforts at Georgetown University to implement a CoCM program in safety net clinics, and partnered with the AIMS Center, the American Psychiatric Association (APA) and the Kennedy Forum to facilitate reimbursement for CoCM. Her work focusing on delivery of effective behavioral health services in primary care settings has expanded to include a focus on measurement based care. Dr. Alter has served in multiple leadership roles in the APA related to quality and health system delivery, including representation to URAC, and co-chair of the CMS/AHIP Behavioral Health Core Quality Measures Collaborative. As well, she is a past president of the Academy of Consult Liasion Psychiatry.

She attended George Washington University Medical School and received psychiatric training at the Mayo Clinic, Cornell University Medical College, and Memorial Sloan Kettering Cancer Center.

Disclosures: Prof. Alter serves in an advisory capacity with, receives consulting fees from, and is an options holder with Kivira Health. She also hold stock in Mindoula Health, and Vitalic Health.

Dr. Laura Vowels, PhD

University of Roehampton

Laura Vowels is an assistant professor and the programme lead for the MSc in Digital Health at University of Roehampton. Her work focuses on relationship functioning, sexual health, psychotherapy and Artificial Intelligence for psychotherapy and risk detection. She is also a practising marriage and family therapist in private practice and has previously she worked as a Psychological Wellbeing Practitioner (PWP) within the NHS Improving Access to Psychological Therapies (IAPT) programme in the UK.

Disclosures: Dr. Laura Vowels is married to the first author on this work, and has received research funding from the Swiss National Science Foundation, Kivira Health, Innovate UK, the ESRC Digital Good Network, the ESRC, InsideOut, and the European Association for Social Psychology.

Dr. Kathryn Erickson-Ridout, MD, PhD

Kaiser Permanente Division of Research

Kathryn Erickson-Ridout is a clinical and translational researcher with expertise in mental health outcomes, health systems research, and population-level analysis. Her work focuses on understanding and improving mental health care delivery within large integrated healthcare systems.

Disclosures: Besides Dr. Erickson-Ridout's position at Kaiser Permanente, she serves in an advisory capacity with and receives consulting fees from Kivira Health.

References

Abi-Dargham, A., Moeller, S. J., Ali, F., DeLorenzo, C., Domschke, K., Horga, G., Jutla, A., et al. (2023). Candidate biomarkers in psychiatric disorders: State of the field. *World Psychiatry*, *22*(2), 236–262. https://doi.org/10.1002/wps.21078

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., Nichols, C. N., Lampropoulos, G. K., Walker, B. S., Cohen, G., & Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist*, *34*(3), 341–382. https://doi.org/10.1177/0011000005285875

Aledavood, T., Torous, J., Triana Hoyos, A. M., Naslund, J. A., Onnela, J.-P., & Keshavan, M. (2019). Smartphone-based tracking of sleep in depression, anxiety, and psychotic disorders. *Current Psychiatry Reports*, *21*(7), 49. https://doi.org/10.1007/s11920-019-1043-y

Alzheimer's Association. (2022). 2022 alzheimer's disease facts and figures. *Alzheimer's & Dementia*, *18*(4), 700–789. https://doi.org/10.1002/alz.12638

American Medical Association. (2024a). CPT code 96127: Brief emotional/behavioral assessment, with scoring and documentation, per standardized instrument [American Medical Association, Chicago, IL].

American Medical Association. (2024b). Psychiatric collaborative care management services CPT code range 99492- 99494 [American Medical Association, Chicago, IL].

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th Text Revision).

American Psychiatric Association. (2023). Resource document on implementation of measurement-based care [American Psychiatric Association Resource Document]. https://www.psychiatry.org/getattachment/3d9484a0-4b8e-4234-bd0d-c35843541fce/Resource-Document-on-Implementation-of-Measurement-Based-Care.pdf

American Psychological Association. (2019). Clinical practice guideline for the treatment of depression across three age cohorts. https://www.apa.org/depression-guideline

Angelakis, I., Gooding, P. A., & Panagioti, M. (2016). Suicidality in body dysmorphic disorder (BDD): A systematic review with meta-analysis. *Clinical Psychology Review*, *49*, 55–66. https://doi.org/10.1016/j.cpr.2016.08.002

Archer, J., Bower, P., Gilbody, S., Lovell, K., Richards, D., Gask, L., Dickens, C., & Coventry, P. (2012). Collaborative care for depression and anxiety problems. *Cochrane Database of Systematic Reviews*, *10*, CD006525. https://doi.org/10.1002/14651858.CD006525.pub2

Arias, D., Saxena, S., & Verguet, S. (2022). Quantifying the global burden of mental disorders and their economic value. *eClinicalMedicine*, *54*, 101675. https://doi.org/10.1016/j.eclinm.2022.101675

Armstrong, T., & Olatunji, B. O. (2012). Eye tracking of attention in the affective disorders: A meta-analytic review and synthesis. *Clinical Psychology Review*, *32*(8), 704–723. https://doi.org/10.1016/j.cpr.2012.09.004

Association, A. C. H. (2023). National college health assessment iii: Reference group executive summary spring 2023. *ACHA*. https://www.acha.org/wp-content/uploads/2024/07/NCHA-III_SPRING_2023_REFERENCE_GROUP_EXECUTIVE_SUMMARY.pdf

Association for University and College Counseling Center Directors. (2023). Aucccd annual survey 2023. https://www.aucccd.org/assets/documents/Survey/2022-2023%20Annual%20Survey%20Report%20Public.pdf

Attridge, M., & Dickens, S. P. (2022). Health and work outcomes of brief counseling from an employee assistance program in vermont: Follow-up survey results, client satisfaction, and estimated cost savings. *SAGE Open*, *12*(1). https://doi.org/10.1177/21582440221087278

Bach, B., & First, M. B. (2018). Application of the ICD-11 classification of personality disorders. *BMC Psychiatry*, *18*(351), 351. https://doi.org/10.1186/s12888-018-1908-3

Bahji, A., Vazquez, G. H., & Zarate, C. A. (2021). Comparative efficacy of racemic ketamine and esketamine for depression: A systematic review and meta-analysis. *Journal of Affective Disorders*, *278*, 542–555. https://doi.org/10.1016/j.jad.2020.09.071

Bandelow, B., Allgulander, C., Baldwin, D. S., Costa, D. L. d. C., Denys, D., Dilbaz, N., Domschke, K., et al. (2023). World Federation of Societies of Biological Psychiatry (wfsbp) and Canadian Anxiety Guidelines/psychdb (CAGP) Guidelines for the pharmacological treatment of anxiety disorders, obsessive-compulsive disorder, and posttraumatic stress disorder. *The World Journal of Biological Psychiatry*, *24*(2), 79–117. https://doi.org/10.1080/15622975.2022.2086295

Baumel, A., Muench, F., Edan, S., & Kane, J. M. (2019). Objective user engagement with mental health apps: Systematic search and panel-based usage analysis. *Journal of Medical Internet Research*, *21*(9). https://doi.org/10.2196/14567

Behave Health. (2024). *Mastering cpt code 96127: A comprehensive guide to brief emotional/behavioral assessment billing* [Accessed 2026-05-14]. https://behavehealth.com/blog/mastering-cpt-code-96127-a-comprehensive-guide-to-brief-emotionalbehavioral-assessment-billing

Belsher, B. E., Smolenski, D. J., Pruitt, L. D., Bush, N. E., Beech, E. H., Workman, D. E., Morgan, R. L., Evatt, D. P., Tucker, J., & Skopp, N. A. (2019). Prediction models for suicide attempts and deaths: A systematic review and simulation. *JAMA Psychiatry*, *76*(6), 642–651. https://doi.org/10.1001/jamapsychiatry.2019.0174

Bhui, K., & Bhugra, D. (2002). Explanatory models for mental distress: Implications for clinical practice and research. *The British Journal of Psychiatry*, *181*(1), 6–7. https://doi.org/10.1192/bjp.181.1.6

Bickerdike, L., Booth, A., Wilson, P. M., Farley, K., & Wright, K. (2017). Social prescribing: Less rhetoric and more reality. a systematic review of the evidence. *BMJ Open*, *7*. https://doi.org/10.1136/bmjopen-2016-013384

Bilder, R. M., & Reise, S. P. (2019). Neuropsychological tests of the future: How do we get there from here? *The Clinical Neuropsychologist*, *33*(2), 220–242. https://doi.org/10.1080/13854046.2018.1521993

Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., et al. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, *582*, 84–88. https://doi.org/10.1038/s41586-020-2314-9

Boudreaux, E. D., Niro, K., Sullivan, A., Rosenbaum, C. D., Allen, M., & Camargo, C. A. (2011). Current practices for mental health follow-up after psychiatric emergency department/psychiatric emergency service visits: A national survey of academic emergency departments. *General Hospital Psychiatry*, *33*(6), 631–633. https://doi.org/10.1016/j.genhosppsych.2011.05.020

Bouzikos, S., Afsharian, A., Dollard, M., & Brecht, O. (2022). Contextualising the effectiveness of an employee assistance program intervention on psychological health: The role of corporate climate. *International Journal of Environmental Research and Public Health*, *19*(9), 5067. https://doi.org/10.3390/ijerph19095067

Bradford, A., Meyer, A. N. D., Khan, S., Giardina, T. D., & Singh, H. (2024). Diagnostic error in mental health: A review. *BMJ Quality & Safety*, *33*(10), 663–672. https://doi.org/10.1136/bmjqs-2023-016996

Brady, K. T., Killeen, T. K., Brewerton, T., & Lucerini, S. (2000). Comorbidity of psychiatric disorders and posttraumatic stress disorder. *Journal of Clinical Psychiatry*, *61*(Suppl 7), 22–32.

Bromet, E., Andrade, L. H., Hwang, I., Sampson, N. A., Alonso, J., de Girolamo, G., et al. (2011). Cross-national epidemiology of dsm-iv major depressive episode. *BMC Medicine*, *9*(90). https://doi.org/10.1186/1741-7015-9-90

Brown, E. L., Raue, P., Halpert, K. D., Adams, S., & Titler, M. G. (2009). Evidence-based guideline: Detection of depression in older adults. *Journal of Gerontological Nursing*, *35*(2), 11–15. https://doi.org/10.3928/00989134-20090201-08

Browne, C. C., Korobanova, D., Chemjong, P., Harris, A. W. F., Glozier, N., Basson, J., Spencer, S.-J., & Dean, K. (2022). Continuity of mental health care during the transition from prison to the community following brief periods of imprisonment. *Frontiers in Psychiatry*, *13*, 934837. https://doi.org/10.3389/fpsyt.2022.934837

Bufano, P., Laurino, M., Said, S., Tognetti, A., & Menicucci, D. (2023). Digital phenotyping for monitoring mental disorders: Systematic review. *J. Med. Internet Res.* https://doi.org/10.2196/46778

Cantor, J., Schuler, M. S., Matthews, S., Kofner, A., Breslau, J., & McBain, R. K. (2024). Availability of mental telehealth services in the us. *JAMA Health Forum*, *5*(2), e235142. https://doi.org/10.1001/jamahealthforum.2023.5142

Carey, M., Jones, K., Meadows, G., Sanson-Fisher, R., D'Este, C., Inder, K., Yoong, S. L., & Russell, G. (2014). Accuracy of general practitioner unassisted detection of depression. *Australian & New Zealand Journal of Psychiatry*, *48*(6), 571–578. https://doi.org/10.1177/0004867413520047

Carhart-Harris, R., & Goodwin, G. M. (2017). The therapeutic potential of psychedelic drugs: Past, present, and future. *Neuropsychopharmacology*, *42*(11), 2105–2113. https://doi.org/10.1038/npp.2017.84

Carlbring, P., Andersson, G., Cuijpers, P., Riper, H., & Hedman-Lagerlöf, E. (2018). Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: An updated systematic review and meta-analysis. *Cognitive Behaviour Therapy*, *47*(1), 1–18. https://doi.org/10.1080/16506073.2017.1401115

Centers for Medicare & Medicaid Services. (2024). Chronic care management services [Accessed 2026-05-14]. https://www.cms.gov/files/document/chroniccaremanagement.pdf

Cernich, A. N., Brennana, D. M., Barker, L. M., & Bleiberg, J. (2007). Sources of error in computerized neuropsychological assessment. *Archives of Clinical Neuropsychology*, *22*, S39–S48. https://doi.org/10.1016/j.acn.2006.10.004

Chan, A., Rahimi-Ardabilli, H., Rogers, W. A., & Coiera, E. (2025). The real-world impact of artificial intelligence ethics frameworks across a decade in healthcare: A scoping review. *Journal of the American Medical Informatics Association*, *32*(11), 1767–1777. https://doi.org/10.1093/jamia/ocaf167

Chang, G., Weiss, A. P., Orav, E. J., Smallwood, J. A., Gonzalez, S., Kosowsky, J. M., & Rauch, S. L. (2012). Bottlenecks in the emergency department: The psychiatric clinicians' perspective. *General Hospital Psychiatry*, *34*, 10–15. https://doi.org/10.1016/j.genhosppsych.2012.03.005

Chisholm, D., Sweeny, K., Sheehan, P., Rasmussen, B., Smit, F., Cuijpers, P., & Saxena, S. (2016). Scaling-up treatment of depression and anxiety: A global return on investment analysis. *The Lancet Psychiatry*, *3*(5), 415–424. https://doi.org/10.1016/S2215-0366(16)30024-4

Cipriani, A., Furukawa, T. A., Salanti, G., et al. (2018). Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: A systematic review and network meta-analysis. *The Lancet*, *391*(10128), 1357–1366. https://doi.org/10.1016/S0140-6736(17)32802-7

Cohn, J. F., Cummins, N., Epps, J., Goecke, R., Joshi, J., & Scherer, S. (2018). Multimodal assessment of depression from behavioral signals. In S. Oviatt, B. Schuller, P. R. Cohen, D. Sonntag, G. Potamianos, & A. Krüger (Eds.), *The handbook of multimodal-multisensor interfaces: Signal processing, architectures, and detection of emotion and cognition* (pp. 375–417, Vol. 2). Association for Computing Machinery. https://doi.org/10.1145/3107990.3108004

Coons, S. J., Gwaltney, C. J., Hays, R. D., et al. (2009). Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome measures: Ispor epro good research practices task force report. *Value in Health*, *12*(4), 419–429. https://doi.org/10.1111/j.1524-4733.2008.00470.x

Correll, C. U., Solmi, M., Croatto, G., Schneider, L. K., Rohani-Montez, S. C., Fairley, L., Smith, N., Bitter, I., Gorwood, P., Taipale, H., & Tiihonen, J. (2022). Mortality in people with schizophrenia: A systematic review and meta-analysis of relative risk and aggravating or attenuating factors. *World Psychiatry*, *21*(2), 248–271. https://doi.org/10.1002/wps.20994

Coulter, C., Baker, K. K., & Margolis, R. L. (2019). Specialized consultation for suspected recent-onset schizophrenia: Diagnostic clarity and the distorting impact of anxiety and reported auditory hallucinations. *Journal of Psychiatric Practice*, *25*(2), 76–81. https://doi.org/10.1097/PRA.0000000000000363

Cuellar, A. K., Johnson, S. L., & Winters, R. (2005). Distinctions between bipolar and unipolar depression. *Clinical Psychology Review*, *25*(3), 307–339. https://doi.org/10.1016/j.cpr.2004.12.002

Cuijpers, P., Cristea, I. A., Karyotaki, E., Reijnders, M., & Huibers, M. J. H. (2016). How effective are cognitive behavior therapies for major depression and anxiety disorders? a meta-analytic update of the evidence. *World Psychiatry*, *15*(3), 245–258. https://doi.org/10.1002/wps.20346

Cuijpers, P., Karyotaki, E., de Wit, L., & Ebert, D. D. (2020). The effects of fifteen evidence-supported therapies for adult depression: A meta-analytic review. *Psychotherapy Research*, *30*(3), 279–293. https://doi.org/10.1080/10503307.2019.1649732

Cuijpers, P., Noma, H., Karyotaki, E., Cipriani, A., & Furukawa, T. A. (2019). Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: A network meta-analysis. *JAMA Psychiatry*, *76*(7), 700–707. https://doi.org/10.1001/jamapsychiatry.2020.0268

Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech Communication*, *71*, 10–49. https://doi.org/10.1016/j.specom.2015.03.004

Daveney, J., Panagioti, M., Waheed, W., & Esmail, A. (2019). Unrecognized bipolar disorder in patients with depression managed in primary care: A systematic review and meta-analysis. *General Hospital Psychiatry*, *58*, 14–21. https://doi.org/10.1016/j.genhosppsych.2019.03.006

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, *141*(4), 858–900. https://doi.org/10.1037/a0038498

Dell'Osso, B., Benatti, B., Buoli, M., Altamura, A. C., Marazziti, D., Hollander, E., & Fineberg, N. (2013). The influence of age at onset and duration of untreated illness on long-term outcome in obsessive-compulsive disorder: A report from the international college of obsessive-compulsive spectrum disorders (icocs). *European Neuropsychopharmacology*, *23*(8), 865–871. https://doi.org/10.1016/j.euroneuro.2013.05.004

Dreisbach, C., Koleck, T. A., Bourne, P. E., & Bakken, S. (2019). A systematic review of natural language processing and text mining of symptoms

from electronic patient-authored text data. *International Journal of Medical Informatics*, *125*, 37–46. https://doi.org/10.1016/j.ijmedinf.2019.02.008

Drug Enforcement Administration & Department of Health and Human Services. (2025). Fourth temporary extension of covid-19 telemedicine flexibilities for prescription of controlled medications.

Edinger, J. D., Arnedt, J. T., Bertisch, S. M., Carney, C. E., Harrington, J. J., Lichstein, K. L., Sateia, M. J., Troxel, W. M., Zhou, E. S., Kazmi, U., Heald, J. L., & Martin, J. L. (2021). Behavioral and psychological treatments for chronic insomnia disorder in adults: An American Academy of Sleep Medicine clinical practice guideline. *Journal of Clinical Sleep Medicine*, *17*(2), 255–262. https://doi.org/10.5664/jcsm.8986

Ein, N., Gervasio, J., St. Cyr, K., Liu, J. J. W., Baker, C., Nazarov, A., & Richardson, J. D. (2024). A rapid review of the barriers and facilitators of mental health service access among veterans and their families. *Frontiers in Health Services*, *4*, 1426202. https://doi.org/10.3389/frhs.2024.1426202

Ellison, W. D., Rosenstein, L. K., Morgan, T. A., & Zimmerman, M. (2018). Community and clinical epidemiology of borderline personality disorder. *Psychiatric Clinics of North America*, *41*(4), 561–573. https://doi.org/10.1016/j.psc.2018.07.008

Etkin, A. (2019). A reckoning and research agenda for neuroimaging in psychiatry. *American Journal of Psychiatry*, *176*(7), 507–511. https://doi.org/10.1176/appi.ajp.2019.19050521

Faraone, S. V., Asherson, P., Banaschewski, T., Coghill, D., Zheng, Y., Biederman, J., et al. (2021). The world federation of adhd international consensus statement: 208 evidence-based conclusions about the disorder. *Neuroscience and Biobehavioral Reviews*, *128*, 789–818. https://doi.org/10.1016/j.neubiorev.2021.01.022

Fayyad, J., Sampson, N. A., Hwang, I., Adamowski, T., Aguilar-Gaxiola, S., Al-Hamzawi, A., Andrade, L. H., et al. (2017). The descriptive epidemiology of dsm-iv adult adhd in the world health organization world mental health surveys. *Attention Deficit and Hyperactivity Disorders*, *9*(1), 47–65. https://doi.org/10.1007/s12402-016-0208-3

Fazel, S., & Seewald, K. (2012). Severe mental illness in 33,588 prisoners worldwide: Systematic review and meta-regression analysis. *The British Journal of Psychiatry*, *200*(5), 364–373. https://doi.org/10.1192/bjp.bp.111.096370

Fernandez, E., Woldgabreal, Y., Day, A., Pham, T., Gleich, B., & Aboujaoude, E. (2021). Live psychotherapy by video versus in-person: A meta-analysis of efficacy and its relationship to types and targets of treatment. *Clinical Psychology & Psychotherapy*, *28*(6), 1535–1549. https://doi.org/10.1002/cpp.2594

First, M., Carter, C., Castellanos, F., Dickstein, D., Drevets, W., Kim, K., Pescosolido, M., et al. (2012). Consensus report of the apa work group on neuroimaging markers of psychiatric disorders. *American Psychiatric Association Resource Document*. https://api.semanticscholar.org/CorpusID:44484439

First, M. B., Williams, J. B. W., Karg, R. S., & Spitzer, R. L. (2016). *Structured clinical interview for dsm-5 disorders, clinician version (scid-5-cv)*. American Psychiatric Association.

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, *3*(4), 456–465. https://doi.org/10.1177/2515245920952393

Fortney, J. C., Unützer, J., Wrenn, G., Pyne, J. M., Smith, G. R., Schoenbaum, M., & Harbin, H. T. (2017). A tipping point for measurement-based care. *Psychiatric Services*, *68*(2), 179–188. https://doi.org/10.1176/appi.ps.201500439

Fung, H. W., Leousi, A., Chau, A. K. C., Sit, K. Y., Lam, C. C., Lay, C. M., Huang, C. H. O., Cheung, C. T. Y., Lam, S. K. K., Wang, E. K. S., & Ho, G. W. K. (2025). Prevalence of icd-11 post-traumatic stress disorder (ptsd) and complex ptsd in the general populations: A systematic review and meta-analysis. *Asian Journal of Psychiatry*, *110*, 104610. https://doi.org/10.1016/j.ajp.2025.104610

Garcia, M. E., Hinton, L., Neuhaus, J., et al. (2022). Equitability of depression screening after implementation of general adult screening in primary care. *JAMA Network Open*, *5*(8), e2227658. https://doi.org/10.1001/jamanetworkopen.2022.27658

GBD 2019 Mental Disorders Collaborators. (2022). Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: A systematic analysis for the global burden of disease study 2019. *The Lancet Psychiatry*, *9*(2), 137–150. https://doi.org/10.1016/S2215-0366(21)00395-3

Gibbons, R. D., Weiss, D. J., Kupfer, D. J., Frank, E., Fagiolini, A., Grochocinski, V. J., Bhaumik, D. K., Stover, A., Bock, R. D., & Immekus, J. C. (2008). Using computerized adaptive testing to reduce the burden of mental health assessment. *Psychiatric Services*, *59*(4), 361–368. https://doi.org/10.1176/ps.2008.59.4.361

Girard, J. M., & Cohn, J. F. (2015). Automated audiovisual depression analysis. *Current Opinion in*

Psychology, *4*, 75–79. https://doi.org/10.1016/j.copsyc.2014.12.010

Gloss, D., Varma, J., Pringsheim, T., & Nuwer, M. R. (2016). Practice advisory: The utility of eeg theta/beta power ratio in adhd diagnosis. *Neurology*, *87*(22), 2375–2379. https://doi.org/10.1212/WNL.0000000000003265

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, *12*(1), 19–30. https://doi.org/10.1037/1040-3590.12.1.19

Gwaltney, C. J., Shields, A. L., & Shiffman, S. (2008). Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: A meta-analytic review. *Value in Health*, *11*(2), 322–333. https://doi.org/10.1111/j.1524-4733.2007.00231.x

Hassan, L., Milton, A., SaAtlawyer, C., Casson, A. J., Torous, J., Davies, A., Ruiz-Yu, B., & Firth, J. (2025). Utility of consumer-grade wearable devices for inferring physical and mental health outcomes in severe mental illness: Systematic review. *JMIR Mental Health*, *12*, e65143. https://doi.org/10.2196/65143

Haven, T. L., & Ioannidis, J. P. A. (2026). Reproducibility failure in biomedical research: Problems and solutions. *Annual Review of Medicine*, *77*(1), 537–549. https://doi.org/10.1146/annurev-med-050124-050859

Health Resources and Services Administration. (2024). Designated Health Professional Shortage Areas statistics: Mental health [Bureau of Health Workforce, HRSA]. https://data.hrsa.gov/topics/health-workforce/shortage-areas

Heim, E., Mewes, R., Abi Ramia, J., Glaesmer, H., Hall, B., Shehadeh, M., Ünlü, B., Kananian, S., Kohrt, B. A., et al. (2021). Reporting cultural adaptation in psychological trials: The RECAPT criteria. *Clinical Psychology in Europe*, *3*, e6351. https://doi.org/10.32872/cpe.6351

Hickey, B. A., Chalmers, T., Newton, P., Lin, C.-T., Sibbritt, D., McLachlan, C. S., Clifton-Bligh, R., Morley, J., & Lal, S. (2021). Smart devices and wearable technologies to detect and monitor mental health conditions and stress: A systematic review. *Sensors*, *21*(10), 3461. https://doi.org/10.3390/s21103461

Hirschfeld, R. M. A., Lewis, L., & Vornik, L. A. (2003). Perceptions and impact of bipolar disorder: How far have we really come? results of the national depressive and manic-depressive association 2000 survey of individuals with bipolar disorder. *Journal of Clinical Psychiatry*, *64*(2), 161–174. https://doi.org/10.4088/JCP.v64n0209

Hoge, C. W., McGurk, D., Thomas, J. L., Cox, A. L., Engel, C. C., & Castro, C. A. (2008). Mild traumatic brain injury in u.s. soldiers returning from iraq. *New England Journal of Medicine*, *358*(5), 453–463. https://doi.org/10.1056/NEJMoa072972

Holmqvist, K., Nyström, M., et al. (2023). *Eye tracking: A comprehensive guide to methods and measures* (2nd ed.). Oxford University Press.

Holt-Lunstad, J. (2024). Social connection as a critical factor for mental and physical health: Evidence, trends, challenges, and future implications. *World Psychiatry*, *23*(3), 312–332. https://doi.org/10.1002/wps.21224

Hunt, G. E., Malhi, G. S., Lai, H. M., & Cleary, M. (2020). Prevalence of comorbid substance use in major depressive disorder in community and clinical settings, 1990–2019: Systematic review and meta-analysis. *Journal of Affective Disorders*, *266*, 288–304. https://doi.org/10.1016/j.jad.2020.01.141

Huynh, P. A., Kindred, R., Perrins, K., de Boer, K., Miles, S., Bates, G., & Nedeljkovic, M. (2025). Prevalence of Complex Post-Traumatic Stress Disorder (CPTSD): A systematic review and meta-analysis. *Psychiatry Research*, *351*, 116586. https://doi.org/10.1016/j.psychres.2025.116586

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research domain criteria (rdoc): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, *167*(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Insel, T. R. (2017). Digital phenotyping: Technology for a new science of behavior. *JAMA*, *318*(13), 1215–1216. https://doi.org/10.1001/jama.2017.11295

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Ioannidis, J. P. A. (2008). Effectiveness of antidepressants: An evidence myth constructed from a thousand randomized trials? *Philosophy, Ethics, and Humanities in Medicine*, *3*(14). https://doi.org/10.1186/1747-5341-3-14

Jacobson, N. C., & Bhattacharya, S. (2022). Digital biomarkers of anxiety disorder symptom changes: Personalized deep learning models using smartphone sensors accurately predict anxiety symptoms from ecological momentary assessments. *Behaviour Research and Therapy*, *149*, 104013. https://doi.org/10.1016/j.brat.2021.104013

Jiang, Z., Luskus, M., Seyedi, S., Giner, E. L., Rad, A. B., Clifford, G. D., Boazak, m., & Cotes,

R. O. (2022). Utilizing computer vision for facial behavior analysis in schizophrenia studies: A systematic review. *PLOS ONE*, *17*(4), e0266828. https://doi.org/10.1371/journal.pone.0266828

Judd, L. L., Akiskal, H. S., Schettler, P. J., Endicott, J., Maser, J., Solomon, D. A., Leon, A. C., et al. (2002). The long-term natural history of the weekly symptomatic status of bipolar i disorder. *Archives of General Psychiatry*, *59*(6), 530–537. https://doi.org/10.1001/archpsyc.59.6.530

Katayama, E. S., Woldesenbet, S., Munir, M. M., Bryan, C. J., Carpenter, K. M., & Pawlik, T. M. (2023). Geospatial analysis of psychiatry workforce distribution and patient travel time reveals disparities in access to mental healthcare. *Psychiatry Research Communications*, *3*(4), 100136. https://doi.org/10.1016/j.psycom.2023.100136

Keepers, G. A., Fochtmann, L. J., Anzia, J. M., Benjamin, S., Lyness, J. M., Mojtabai, R., Servis, M., Walaszek, A., Buckley, P., Lenzenweger, M. F., Young, A. S., Degenhardt, A., & Hong, S.-H. (2020). The American Psychiatric Association practice guideline for the treatment of patients with schizophrenia. *American Journal of Psychiatry*, *177*(9), 868–872. https://doi.org/10.1176/appi.ajp.2020.177901

Kelly, T. M., & Daley, D. C. (2013). Integrated treatment of substance use and psychiatric disorders. *Social Work in Public Health*, *28*(3-4), 388–406. https://doi.org/10.1080/19371918.2013.774673

Keramatian, K., Chithra, N. K., & Yatham, L. N. (2023). The CANMAT and ISBD guidelines for the treatment of bipolar disorder: Summary and a 2023 update of evidence. *Focus*, *21*(4), 344–353. https://doi.org/10.1176/appi.focus.20230009

Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R., & Walters, E. E. (2005a). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*(6), 617–627. https://doi.org/10.1001/archpsyc.62.6.617

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005b). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*(6), 593–602. https://doi.org/10.1001/archpsyc.62.6.593

Kinnaird, E., Stewart, C., & Tchanturia, K. (2019). Investigating alexithymia in autism: A systematic review and meta-analysis. *European Psychiatry*, *55*, 80–89. https://doi.org/10.1016/j.eurpsy.2018.09.004

Knapp, M., & Wong, G. (2020). Economics and mental health: The current scenario. *World Psychiatry*, *19*(1), 3–14. https://doi.org/10.1002/wps.20692

Knutson, K. H., Wennberg, D., & Rajkumar, R. (2021). Driving access and quality: A shift to value-based behavioral health care. *Psychiatric Services*, *72*(8), 942–948. https://doi.org/10.1176/appi.ps.202000386

Koenen, K. C., Ratanatharathorn, A., Ng, L., McLaughlin, K. A., Bromet, E. J., Stein, D. J., Karam, E. G., et al. (2017). Posttraumatic stress disorder in the world mental health surveys. *Psychological Medicine*, *47*(13), 2260–2274. https://doi.org/10.1017/S0033291717000708

Kohn, R., Saxena, S., Levav, I., & Saraceno, B. (2004). The treatment gap in mental health care. *Bulletin of the World Health Organization*, *82*(11), 858–866.

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., ... Zimmerman, M. (2017). The hierarchical taxonomy of psychopathology (hitop): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454–477. https://doi.org/10.1037/abn0000258

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Krug, I., Liu, S., Portingale, J., Croce, S., Dar, B., Obleada, K., Satheesh, V., Wong, M., & Fuller-Tyszkiewicz, M. (2025). A meta-analysis of mortality rates in eating disorders: An update of the literature from 2010 to 2024. *Clinical Psychology Review*, *116*, 102547. https://doi.org/10.1016/j.cpr.2025.102547

Lai, M.-C., Kassee, C., Besney, R., Bonato, S., Hull, L., Mandy, W., Szatmari, P., & Ameis, S. H. (2019). Prevalence of co-occurring mental health diagnoses in the autism population: A systematic review and meta-analysis. *The Lancet Psychiatry*, *6*(10), 819–829. https://doi.org/10.1016/S2215-0366(19)30289-5

Larsen, M. E., Huckvale, K., Nicholas, J., Torous, J., Birrell, L., Li, E., & Reda, B. (2019). Using science to sell apps: Evaluation of mental health app store quality claims. *npj Digital Medicine*, *2*, 18. https://doi.org/10.1038/s41746-019-0093-1

Le Glaz, A., Haralambous, Y., Kim-Dufor, D.-H., Lenca, P., Billot, R., Ryan, T. C., Marsh, J., DeVylder,

J., Walter, M., Berrouiguet, S., & Lemey, C. (2021). Machine learning and natural language processing in mental health: Systematic review. *Journal of Medical Internet Research*, *23*(5), e15708. https://doi.org/10.2196/15708

Lecomte, T., Potvin, S., Corbière, M., Guay, S., Samson, C., Cloutier, B., Francoeur, A., Pennou, A., & Khazaal, Y. (2020). Mobile apps for mental health issues: Meta-review of meta-analyses. *JMIR mHealth and uHealth*, *8*(5), e17458. https://doi.org/10.2196/17458

Leichsenring, F., Fonagy, P., Heim, N., Kernberg, O. F., Leweke, F., Luyten, P., Salzer, S., Spitzer, C., & Steinert, C. (2024). Borderline personality disorder: A comprehensive review of diagnosis and clinical presentation, etiology, treatment, and current controversies. *World Psychiatry*, *23*(1), 4–25. https://doi.org/10.1002/wps.21156

Leichsenring, F., Steinert, C., Rabung, S., & Ioannidis, J. P. A. (2022). The efficacy of psychotherapies and pharmacotherapies for mental disorders in adults: An umbrella review and meta-analytic evaluation of recent meta-analyses. *World Psychiatry*, *21*(1), 133–145. https://doi.org/10.1002/wps.20941

Lew, H. L., Otis, J. D., Tun, C., Kerns, R. D., Clark, M. E., & Cifu, D. X. (2009). Prevalence of chronic pain, posttraumatic stress disorder, and persistent postconcussive symptoms in OIF/OEF veterans: Polytrauma clinical triad. *Journal of Rehabilitation Research and Development*, *46*(6), 697–702. https://doi.org/10.1682/JRRD.2009.01.0006

Lewis, C. C., Boyd, M., Puspitasari, A., Navarro, E., Howard, J., Kassab, H., Hoffman, M., Scott, K., Lyon, A., Douglas, S., Simon, G., & Kroenke, K. (2019). Implementing measurement-based care in behavioral health: A review. *JAMA Psychiatry*, *76*(3), 324–335. https://doi.org/10.1001/jamapsychiatry.2018.3329

Lish, J. D., Dime-Meenan, S., Whybrow, P. C., Price, R. A., & Hirschfeld, R. M. (1994). The national depressive and manic-depressive association (NDMDA) survey of bipolar members. *Journal of Affective Disorders*, *31*(4), 281–294. https://doi.org/10.1016/0165-0327(94)90104-X

Low, D. M., Bentley, K. H., & Ghosh, S. S. (2020). Automated assessment of psychiatric disorders using speech: A systematic review. *Laryngoscope Investigative Otolaryngology*, *5*(1), 96–116. https://doi.org/10.1002/lio2.354

Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., & Munafò, M. R. (2016). Gamification of cognitive assessment and cognitive training: A systematic review of applications and efficacy. *JMIR Serious Games*, *4*(2), e11. https://doi.org/10.2196/games.5888

Malgaroli, M., Hull, T. D., Zech, J. M., & Althoff, T. (2023). Natural language processing for mental health interventions: A systematic review and research framework. *Translational Psychiatry*, *13*, 309. https://doi.org/10.1038/s41398-023-02592-2

Malhi, G. S., & Mann, J. J. (2018). Depression. *The Lancet*, *392*(10161), 2299–2312. https://doi.org/10.1016/S0140-6736(18)31948-2

Martinez-Martin, N., Insel, T. R., Dagum, P., Greely, H. T., & Cho, M. K. (2018). Data mining for health: Staking out the ethical territory of digital phenotyping. *npj Digital Medicine*, 68. https://doi.org/10.1038/s41746-018-0075-8

Martinez-Martin, N., Kreitmair, K., & Cho, M. K. (2021). Ethical development of digital phenotyping tools for mental health applications: Delphi study. *JMIR mHealth and uHealth*, *9*(7). https://doi.org/10.2196/27343

McBain, R. K., Schuler, M. S., Breslau, J., Kofner, A., Wang, L., & Cantor, J. H. (2024). Telehealth availability for mental health care during and after the covid-19 public health emergency. *JAMA Network Open*, *7*(7), e2420853. https://doi.org/10.1001/jamanetworkopen.2024.20853

McCabe, M. P., Sharlip, I. D., Atalla, E., et al. (2016a). Definitions of sexual dysfunctions in women and men: A consensus statement from the fourth international consultation on sexual medicine 2015. *The Journal of Sexual Medicine*, *13*(2), 135–143. https://doi.org/10.1016/j.jsxm.2015.12.019

McCabe, M. P., Sharlip, I. D., Lewis, R., Atalla, E., Balon, R., Fisher, A. D., Laumann, E., Lee, S. W., & Segraves, R. T. (2016b). Incidence and prevalence of sexual dysfunction in women and men: A consensus statement from the fourth international consultation on sexual medicine 2015. *The Journal of Sexual Medicine*, *13*(2), 144–152. https://doi.org/10.1016/j.jsxm.2015.12.034

Merikangas, K. R., Jin, R., He, J.-P., Kessler, R. C., Lee, S., Sampson, N. A., Viana, M. C., et al. (2011). Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. *Archives of General Psychiatry*, *68*(3), 241–251. https://doi.org/10.1001/archgenpsychiatry.2011.12

Miller, I. W., Camargo, C. A., Arias, S. A., Sullivan, A. F., Allen, M. H., Goldstein, A. B., Manton, A. P., Espinola, J. A., Jones, R., Hasegawa, K., & Boudreaux, E. D. (2017). Suicide prevention in an emergency department population: The ed-safe study. *JAMA Psychiatry*, *74*(6), 563–

570. https://doi.org/10.1001/jamapsychiatry.2017.0678

Mohr, D. C., Zhang, M., & Schueller, S. M. (2017). Personal sensing: Understanding mental health using ubiquitous sensors and machine learning. *Annual Review of Clinical Psychology*, *13*, 23–47. https://doi.org/10.1146/annurev-clinpsy-032816-044949

Moura, I., Teles, A., Viana, D., Marques, J., Coutinho, L., & Silva, F. (2023). Digital phenotyping of mental health using multimodal sensing of multiple situations of interest: A systematic literature review. *Journal of Biomedical Informatics*, *138*.

Muehlhausen, W., Doll, H., Quadri, N., Fordham, B., O'Donohoe, P., Dogar, N., & Wild, D. J. (2015). Equivalence of electronic and paper administration of patient-reported outcome measures: A systematic review and meta-analysis of studies conducted between 2007 and 2013. *Health and Quality of Life Outcomes*, *13*, 167. https://doi.org/10.1186/s12955-015-0362-x

Muench, U., Jura, M., Parks Thomas, C., Perloff, J., & Spetz, J. (2022). Rural-urban prescribing patterns by primary care and behavioral health providers in older adults with serious mental illness. *BMC Health Services Research*, *22*(1), 1440. https://doi.org/10.1186/s12913-022-08813-6

Munder, T., Brütsch, O., Leonhart, R., Gerger, H., & Barth, J. (2013). Researcher allegiance in psychotherapy outcome research: An overview of reviews. *Clinical Psychology Review*, *33*(4), 501–511. https://doi.org/10.1016/j.cpr.2013.02.002

Myin-Germeys, I., Kasanova, Z., Vaessen, T., Vachon, H., Kirtley, O., Viechtbauer, W., & Reininghaus, U. (2018). Experience sampling methodology in mental health research: New insights and technical developments. *World Psychiatry*, *17*(2), 123–132. https://doi.org/10.1002/wps.20513

National Institute for Health and Care Excellence. (2005, November). Obsessive-compulsive disorder and body dysmorphic disorder: Treatment. *National Institute for HealthCare Excellence*. https://www.nice.org.uk/guidance/cg31

National Institute for Health and Care Excellence. (2011a). Alcohol-use disorders. *National Institute for HealthCare Excellence*.

National Institute for Health and Care Excellence. (2011b). Generalised anxiety disorder and panic disorder in adults. *National Institute for HealthCare Excellence*.

National Institute for Health and Care Excellence. (2014). *Psychosis and schizophrenia in adults: Prevention and management (NICE guideline CG178)* (tech. rep.). National Institute for Health and Care Excellence. London. https://www.nice.org.uk/guidance/cg178

National Institute for Health and Care Excellence. (2017). Eating disorders. *National Institute for HealthCare Excellence*.

National Institute for Health and Care Excellence. (2018). Attention deficit hyperactivity disorder. *National Institute for HealthCare Excellence*.

National Institute for Health and Care Excellence. (2021). Autism spectrum disorder in adults. *National Institute for HealthCare Excellence*.

National Institute for Health and Care Excellence. (2022a). Depression in adults: Treatment and management [NICE Guideline NG222]. https://www.nice.org.uk/guidance/ng222

National Institute for Health and Care Excellence. (2022b). Self-harm. *National Institute for HealthCare Excellence*.

National Institute on Aging. (2023). Assessing cognitive impairment in older patients. https://www.nia.nih.gov/health/health-care-professionals-information/assessing-cognitive-impairment-older-patients

Newson, J. J., & Thiagarajan, T. C. (2019). EEG frequency bands in psychiatric disorders: A review of resting state studies. *Frontiers in Human Neuroscience*, *12*, 521. https://doi.org/10.3389/fnhum.2018.00521

Nordstrom, K., Berlin, J. S., Nash, S. S., Shah, S. B., Schmelzer, N. A., & Worley, L. L. M. (2019). Boarding of mentally ill patients in emergency departments: American psychiatric association resource document. *Western Journal of Emergency Medicine*, *20*(5), 690–695. https://doi.org/10.5811/westjem.2019.6.42422

O'Donnell, M. L., Agathos, J. A., Metcalf, O., Gibson, K., & Lau, W. (2019). Adjustment disorder: Current developments and future directions. *International Journal of Environmental Research and Public Health*, *16*(14), 2537. https://doi.org/10.3390/ijerph16142537

O'Donohoe, P., Reasner, D. S., Kovacs, S. M., et al. (2023). Updated recommendations on evidence needed to support measurement comparability among modes of data collection for patient-reported outcome measures: A good practices report of an ispor task force. *Value Health*, *26*(5), 623–633.

O'Neil, M. E., Cheney, T. P., Hart, E. L., Holmes, R. S., Blazina, I., Clauss, K., Yu, Y., Fu, R., & Chou, R. (2024). *Pharmacologic and nonpharmacologic treatments for posttraumatic stress disorder: 2024 update of the evidence base for the PTSD trials standardized data repository* (tech. rep. No. AHRQ Publication No. 25-EHC011). Agency for Healthcare Research and Quality.

Rockville, MD. https://doi.org/10.23970/AHRQEPCPTSD2024

Onnela, J.-P., & Rauch, S. L. (2016). Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology*, *41*(7), 1691–1696. https://doi.org/10.1038/npp.2016.7

Organisation for Economic Co-operation and Development. (2018). *Health at a glance: Europe 2018: State of health in the eu cycle* (Accessed 2026-05-14). OECD Publishing. Paris, France. https://doi.org/10.1787/health_glance_eur-2018-en

Organisation for Economic Co-operation and Development. (2026). *The economic case for preventing mental ill health* (Accessed 2026-05-14). OECD Publishing. Paris, France. https://www.oecd.org/en/publications/the-economic-case-for-preventing-mental-ill-health_16668f16-en/full-report.html

Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *Journal of Strategic Information Systems*, *34*(2), 101885. https://doi.org/10.1016/j.jsis.2024.101885

Paris, J., & Black, D. W. (2015). Borderline personality disorder and bipolar disorder: What is the difference and why does it matter? *Journal of Nervous and Mental Disease*, *203*(1), 3–7. https://doi.org/10.1097/NMD.0000000000000225

Parish, S. J., Simon, J. A., Davis, S. R., Giraldi, A., Goldstein, I., Goldstein, S. W., Kim, N. N., Kingsberg, S. A., Morgentaler, A., Nappi, R. E., Park, K., Stuenkel, C. A., Traish, A. M., & Vignozzi, L. (2021). International society for the study of women's sexual health clinical practice guideline for the use of systemic testosterone for hypoactive sexual desire disorder in women. *The Journal of Sexual Medicine*, *18*(5), 849–867. https://doi.org/10.1016/j.jsxm.2020.10.009

Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., Chisholm, D., Collins, P. Y., Cooper, J. L., et al. (2018). The Lancet Commission on global mental health and sustainable development. *The Lancet*, *392*(10157), 1553–1598. https://doi.org/10.1016/S0140-6736(18)31612-X

Peltola, M. E., Leitinger, M., Halford, J. J., et al. (2023). Routine and sleep eeg: Minimum recording standards of the international federation of clinical neurophysiology and the international league against epilepsy. *Clinical Neurophysiology*, *64*. https://doi.org/10.1111/epi.17448

Peritogiannis, V., Ninou, A., & Samakouri, M. (2022). Mortality in schizophrenia-spectrum disorders: Recent advances in understanding and management. *Healthcare*, *10*(12), 2366. https://doi.org/10.3390/healthcare10122366

Polanczyk, G., de Lima, M. S., Horta, B. L., Biederman, J., & Rohde, L. A. (2007). The worldwide prevalence of adhd: A systematic review and metaregression analysis. *American Journal of Psychiatry*, *164*(6), 942–948. https://doi.org/10.1176/ajp.2007.164.6.942

Prevounce Health. (2026). *Rules for cpt 99490 and the other chronic care management codes* [Accessed 2026-05-14]. https://blog.prevounce.com/rules-for-cpt-99490-and-the-other-chronic-care-management-codes

Regier, D. A., Narrow, W. E., Clarke, D. E., Kraemer, H. C., Kuramoto, S. J., Kuhl, E. A., & Kupfer, D. J. (2013). Dsm-5 field trials in the united states and canada, part ii: Test-retest reliability of selected categorical diagnoses. *American Journal of Psychiatry*, *170*(1), 59–70. https://doi.org/10.1176/appi.ajp.2012.12070999

Riemann, D., Baglioni, C., Bassetti, C., et al. (2017). European guideline for the diagnosis and treatment of insomnia. *Journal of Sleep Research*, *26*(6), 675–700. https://doi.org/10.1111/jsr.12594

Rong, J. e. a. (2025). Global, regional and national burden of depressive disorders and attributable risk factors from 1990 to 2021: Results from the 2021 global burden of disease study. *The British Journal of Psychiatry*. https://doi.org/10.1192/bjp.2024.266

Rural Health Information Hub. (2025). *Chronic care management* [Accessed 2026-05-14]. https://www.ruralhealthinfo.org/care-management/chronic-care-management

Ruscio, A. M., Stein, D. J., Chiu, W. T., & Kessler, R. C. (2010). The epidemiology of obsessive-compulsive disorder in the national comorbidity survey replication. *Molecular Psychiatry*, *15*(1), 53–63. https://doi.org/10.1038/mp.2008.94

Santomauro, D. F., Purcell, C., Whiteford, H. A., Ferrari, A. J., & Vos, T. (2023). Grading disorder severity and averted burden by access to treatment within the gbd framework: A case study with anxiety disorders. *The Lancet Psychiatry*, *10*(4), 272–281. https://doi.org/10.1016/S2215-0366(23)00037-8

Sateia, M. J., Buysse, D. J., Krystal, A. D., Neubauer, D. N., & Heald, J. L. (2017). Clinical practice guideline for the pharmacologic treatment of chronic insomnia in adults: An American Academy of Sleep Medicine clinical practice guideline. *Journal of Clinical Sleep Medicine*, *13*(2), 307–349. https://doi.org/10.5664/jcsm.6470

Schnack, H. G., & Kahn, R. S. (2016). Detecting neuroimaging biomarkers for psychiatric disorders:

Sample size matters. *Frontiers in Psychiatry*, *7*, 50. https://doi.org/10.3389/fpsyt.2016.00050

Schnurr, P. P., Hamblen, J. L., Wolf, J., Coller, R., Collie, C., Fuller, M. A., Holtzheimer, P. E., Kelly, U., Lang, A. J., McGraw, K., Morganstein, J. C., Norman, S. B., Papke, K., Petrakis, I., Riggs, D., Sall, J. A., Shiner, B., Wiechers, I., & Kelber, M. S. (2024). The management of posttraumatic stress disorder and acute stress disorder: Synopsis of the 2023 U.S. Department of Veterans Affairs and U.S. Department of Defense Clinical Practice Guideline. *Annals of Internal Medicine*, *177*(3), 363–374. https://doi.org/10.7326/M23-2757

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., et al. (1998). The mini-international neuropsychiatric interview (m.i.n.i.): The development and validation of a structured diagnostic psychiatric interview for dsm-iv and icd-10. *Journal of Clinical Psychiatry*, *59*(20), 22–33.

Shic, F. (2016). Eye tracking as a behavioral biomarker for psychiatric conditions: The road ahead. *J Am Acad Child Adolesc Psychiatry*, *55*, 267–268. https://doi.org/10.1016/j.jaac.2016.02.002

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, *4*, 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Simmatis, L., Russo, E. E., Geraci, J., Harmsen, I. E., & Samuel, N. (2023). Technical and clinical considerations for electroencephalography-based biomarkers in major depressive disorder. *Npj Ment Health Res*, *2*. https://doi.org/10.1038/s44184-023-00038-7

Simpson, A. I. F., Gerritsen, C., Maheandiran, M., Adamo, V., Vogel, T., Fulham, L., Kitt, T., Forrester, A., & Jones, R. M. (2022). A systematic review of reviews of correctional mental health services using the stair framework. *Frontiers in Psychiatry*, *12*, 747202. https://doi.org/10.3389/fpsyt.2021.747202

Singh, B., Olds, T., Curtis, R., et al. (2023). Effectiveness of physical activity interventions for improving depression, anxiety and distress: An overview of systematic reviews. *British Journal of Sports Medicine*, *57*, 1203–1209. https://doi.org/10.1136/bjsports-2022-106195

Smith, D. J., & Craddock, N. (2011). Unipolar and bipolar depression: Different or the same? *The British Journal of Psychiatry*, *199*(4), 272–274. https://doi.org/10.1192/bjp.bp.111.092726

Song, P., Zha, M., Yang, Q., Zhang, Y., Li, X., & Rudan, I. (2021). The prevalence of adult attention-deficit hyperactivity disorder: A global systematic review and meta-analysis. *Journal of Global Health*, *11*, 04009. https://doi.org/10.7189/jogh.11.04009

Spasic, I., & Nenadic, G. (2020). Clinical text data in machine learning: Systematic review. *JMIR Medical Informatics*, *8*(3), e17984. https://doi.org/10.2196/17984

Stone, A. A., Schneider, S., & Smyth, J. M. (2023). Evaluation of pressing issues in ecological momentary assessment. *Annual Review of Clinical Psychology*, *19*, 107–131. https://doi.org/10.1146/annurev-clinpsy-080921-083128

Substance Abuse and Mental Health Services Administration. (2010). Integrated treatment for co-occurring disorders evidence-based practices (ebp) kit. https://library.samhsa.gov/product/integrated-treatment-co-occurring-disorders-evidence-based-practices-ebp-kit/sma08-4366

Sweetman, J., Knapp, P., Varley, D., Woodhouse, R., & McMillan, D. (2021). Barriers to attending initial psychological therapy service appointments for common mental health problems: A mixed-methods systematic review. *Journal of Affective Disorders*, *284*, 44–63. https://doi.org/10.1016/j.jad.2021.01.089

Tatum, W. O., Selioutski, O., Ochoa, J., Clary, H. M., Cheek, J., Drislane, F., & Tsuchida, T. N. (2016). American clinical neurophysiology society guideline 7: Guidelines for eeg reporting. *Journal of Clinical Neurophysiology*, *33*(4), 328–332. https://doi.org/10.1097/WNP.0000000000000319

Thornicroft, G., Chatterji, S., Evans-Lacko, S., Gruber, M., Sampson, N., Aguilar-Gaxiola, S., Al-Hamzawi, A., Alonso, J., Andrade, L., Borges, G., Bruffaerts, R., Bunting, B., Caldas de Almeida, J. M., Florescu, S., de Girolamo, G., Gureje, O., Haro, J. M., He, Y., Hinkov, H., … Kessler, R. C. (2017). Undertreatment of people with major depressive disorder in 21 countries. *The British Journal of Psychiatry*, *210*(2), 119–124. https://doi.org/10.1192/bjp.bp.116.188078

Torous, J., Linardon, J., Goldberg, S. B., Sun, S., Bell, I., Nicholas, J., Hassan, L., Hua, Y., Milton, A., & Firth, J. (2025). The evolving field of digital mental health: Current evidence and implementation issues for smartphone apps, generative artificial intelligence, and virtual reality. *World Psychiatry*, *24*(2), 156–174. https://doi.org/10.1002/wps.21299

Torous, J., Myrick, K. J., Rauseo-Ricupero, N., & Firth, J. (2020). Digital mental health and covid-19: Using technology today to accelerate the curve on access and quality tomorrow. *JMIR Mental Health*, *7*(3), e18848. https://doi.org/10.2196/18848

Torous, J., Bucci, S., Bell, I. H., et al. (2021). The growing field of digital psychiatry: Current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry*, *20*(3), 318–335. https://doi.org/10.1002/wps.20883

Torous, J., Nicholas, J., Larsen, M. E., Firth, J., & Christensen, H. (2018). Clinical review of user engagement with mental health smartphone apps: Evidence, theory and improvements. *Evidence-Based Mental Health*, *21*(3), 116–119. https://doi.org/10.1136/eb-2018-102891

Torous, J., & Roberts, L. W. (2017). Needed innovation in digital health and smartphone applications for mental health: Transparency and trust. *JAMA Psychiatry*, *74*(5), 437–438. https://doi.org/10.1001/jamapsychiatry.2017.0262

Trull, T. J., & Ebner-Priemer, U. (2013). Ambulatory assessment. *Annual Review of Clinical Psychology*, *9*, 151–176. https://doi.org/10.1146/annurev-clinpsy-050212-185510

Turner, B. O., Paul, E. J., Miller, M. B., & Barbey, A. K. (2018). Small sample sizes reduce the replicability of task-based fmri studies. *Communications Biology*, *1*, 62. https://doi.org/10.1038/s42003-018-0073-z

Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A., & Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *New England Journal of Medicine*, *358*(3), 252–260. https://doi.org/10.1056/NEJMsa065779

United Nations Office on Drugs and Crime. (2024). *World drug report 2024*. United Nations. https://www.unodc.org/unodc/en/data-and-analysis/world-drug-report-2024.html

Unützer, J., Katon, W. J., Fan, M. Y., Schoenbaum, M. C., Lin, E. H. B., Della Penna, R. D., & Powers, D. (2008). Long-term cost effects of collaborative care for late-life depression. *American Journal of Managed Care*, *14*(2), 95–100.

Unützer, J., Katon, W., Callahan, C. M., Williams, J. W., Hunkeler, E., Harpole, L., et al. (2002). Collaborative care management of late-life depression in the primary care setting: A randomized controlled trial. *JAMA*, *288*(22), 2836–2845. https://doi.org/10.1001/jama.288.22.2836

US Preventive Services Task Force. (2020). Screening for cognitive impairment in older adults: Us preventive services task force recommendation statement. *JAMA*, *323*(8), 757–763. https://doi.org/10.1001/jama.2020.0435

US Preventive Services Task Force. (2023). Screening for depression and suicide risk in adults: Us preventive services task force recommendation statement. *JAMA*, *329*(23), 2057–2067. https://doi.org/10.1001/jama.2023.9297

van Hoeken, D., & Hoek, H. W. (2020). Review of the burden of eating disorders: Mortality, disability, costs, quality of life, and family burden. *Current Opinion in Psychiatry*, *33*(6), 521–527. https://doi.org/10.1097/YCO.0000000000000641

van Straten, A., Hill, J., Richards, D. A., & Cuijpers, P. (2015). Stepped care treatment delivery for depression: A systematic review and meta-analysis. *Psychological Medicine*, *45*(2), 231–246. https://doi.org/10.1017/S0033291714000701

Veale, D., Gledhill, L. J., Christodoulou, P., & Hodsoll, J. (2016). Body dysmorphic disorder in different settings: A systematic review and estimated weighted prevalence. *Body Image*, *18*, 168–186. https://doi.org/10.1016/j.bodyim.2016.07.003

Vermani, M., Marcus, M., & Katzman, M. A. (2011). Rates of detection of mood and anxiety disorders in primary care: A descriptive, cross-sectional study. *Primary Care Companion for CNS Disorders*, *13*(2), PCC.10m01013. https://doi.org/10.4088/PCC.10m01013

Volicer, L., Frijters, D., & van der Steen, J. T. (2011). Underdiagnosis and undertreatment of depression in nursing home residents. *European Geriatric Medicine*, *2*(6). https://doi.org/10.1016/j.eurger.2011.08.001

Vowels, M. J. (2023). Misspecification and unreliable interpretations in psychology and social science. *Psychological Methods*, *28*(3), 507–526. https://doi.org/10.1037/met0000429

Waheed, A., Afridi, A. K., Rana, M., Arif, M., Barrera, T., Patel, F., Khan, M. N., & Azhar, E. (2024). Knowledge and behavior of primary care physicians regarding utilization of standardized tools in screening and assessment of anxiety, depression, and mood disorders at a large integrated health system. *Journal of Primary Care & Community Health*, *15*, 21501319231224711. https://doi.org/10.1177/21501319231224711

Wang, P. S., Berglund, P., Olfson, M., Pincus, H. A., Wells, K. B., & Kessler, R. C. (2005). Failure and delay in initial treatment contact after first onset of mental disorders in the national comorbidity survey replication. *Archives of General Psychiatry*, *62*(6), 603–613. https://doi.org/10.1001/archpsyc.62.6.603

Wardenaar, K. J., Lim, C. C. W., Al-Hamzawi, A. O., Alonso, J., Andrade, L. H., Benjet, C., Bunting, B., de Girolamo, G., Demyttenaere, K., Florescu, S. E., Gureje, O., Hisateru, T., et al. (2017). The cross-national epidemiology of specific phobia in the World Mental Health Surveys. *Psychological Medicine*, *47*(10), 1744–1760. https://doi.org/10.1017/S0033291717000174

Whisman, M. A. (2013). Relationship discord and the prevalence, incidence, and treatment of psychopathology. *Journal of Social and Personal Relationships*, *30*(2), 163–170. https://doi.org/10.1177/0265407512455269

Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., Charlson, F. J., Norman, R. E., Flaxman, A. D., Johns,

N., Burstein, R., Murray, C. J. L., & Vos, T. (2013). Global burden of disease attributable to mental and substance use disorders: Findings from the global burden of disease study 2010. *The Lancet*, *382*(9904), 1575–1586. https://doi.org/10.1016/S0140-6736(13)61611-6

Widge, A. S., Bilge, M. T., Montana, R., Chang, W., Rodriguez, C. I., Deckersbach, T., Carpenter, L. L., Kalin, N. H., & Nemeroff, C. B. (2019). Electroencephalographic biomarkers for treatment response prediction in major depressive illness: A meta-analysis. *American Journal of Psychiatry*, *176*(1), 44–56. https://doi.org/10.1176/appi.ajp.2018.17121358

World Health Organization. (2021a). *Ethics and governance of artificial intelligence for health*. https://www.who.int/publications/i/item/9789240029200

World Health Organization. (2021b). Mental health atlas 2020. https://www.who.int/publications/i/item/9789240036703

World Health Organization. (2022a). *International classification of diseases for mortality and morbidity statistics (11th revision)*. https://icd.who.int/browse11

World Health Organization. (2022b). *Mental health at work* [Fact sheet. Accessed 2026-05-14]. https://www.who.int/news-room/fact-sheets/detail/mental-health-at-work

World Health Organization. (2023, September). Anxiety disorders [WHO Fact Sheet]. https://www.who.int/news-room/fact-sheets/detail/anxiety-disorders

World Health Organization. (2025a). Schizophrenia [Accessed 2026-05-15]. https://www.who.int/news-room/fact-sheets/detail/schizophrenia

World Health Organization. (2025b). *World mental health today: Latest data* (tech. rep.). World Health Organization. https://www.who.int/publications/i/item/9789240113817

Wright, J. H., & Caudill, R. (2020). Remote treatment delivery in response to the covid-19 pandemic. *Psychotherapy and Psychosomatics*, *89*(3), 130–132. https://doi.org/10.1159/000507376

Yakovenko, I., Mukaneza, Y., Germé, K., Belliveau, J., Fraleigh, R., Bach, P., Poulin, G., Selby, P., Goyer, M.-È., Brothers, T. D., Rehm, J., Hodgins, D. C., Stewart, S. H., Wood, E., & Bruneau, J. (2024). Management of opioid use disorder: 2024 update to the national clinical practice guideline [Canadian Research Initiative in Substance Matters guideline development team]. *CMAJ*, *196*(38), E1280–E1290. https://doi.org/10.1503/cmaj.241173

Yesavage, J. A., & Sheikh, J. I. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clinical Gerontologist*, *5*(1–2), 165–173. https://doi.org/10.1300/J018v05n01_09

Zeidan, J., Fombonne, E., Scorah, J., et al. (2022). Global prevalence of autism: A systematic review update. *Autism Research*, *15*(5), 778–790. https://doi.org/10.1002/aur.2696

Zhang, T., Schoene, A. M., Ji, S., & Ananiadou, S. (2022). Natural language processing applied to mental illness detection: A narrative review. *npj Digital Medicine*, *5*, 46. https://doi.org/10.1038/s41746-022-00589-7

Zhu, M., Hong, R. H., Yang, T., Yang, X., Wang, X., Liu, J., Murphy, J. K., Michalak, E. E., Wang, Z., Yatham, L. N., Chen, J., & Lam, R. W. (2021). The efficacy of measurement-based care for depressive disorders: Systematic review and meta-analysis of randomized controlled trials. *Journal of Clinical Psychiatry*, *82*(5), 21r14034. https://doi.org/10.4088/JCP.21r14034

A Clinical Scope of Mental Health Conditions and Presentations

This section highlights the most common clinical presentations encountered across care settings. For organisational clarity, we adopt the DSM-5-TR diagnostic framework as the primary nosological reference, consistent with its widespread use in clinical practice, research, reimbursement, and regulatory contexts. However, we also acknowledge that the DSM-5 has been subject to some criticism (e.g., inter-rater reliability for many diagnoses remains modest Regier et al., 2013).

The clinical scope is deliberately broad and encompasses not only formal diagnostic categories but also subthreshold presentations, transdiagnostic constructs, and functional domains that are clinically significant regardless of whether they occur within a single or multiple diagnostic entities. Mental health conditions are highly comorbid, frequently co-occuring with substance use and medical conditions, and are present and evolve across the lifespan. Any standard limited to one or two disorders would reproduce the fragmentation that characterises current care. The following summary illustrates the prevalence and burden of the conditions and presentations in scope.

Depressive disorders are among the leading causes of disability worldwide. The Global Burden of Disease (Rong, 2025) study estimated that depressive disorders accounted for approximately 56.3 million disability-adjusted life years (DALYs) globally in 2021 (Rong, 2025). Twelve-month prevalence estimates of major depressive episode in high-income countries are approximately 5.5%, with lifetime prevalence estimates around 14.6% (Bromet et al., 2011). Major depressive episodes occur in both unipolar major depressive disorder and bipolar disorder; however, substantial evidence suggests that bipolar depression and unipolar depression differ meaningfully in clinical course, family history, neurobiology, and treatment response, raising longstanding concerns regarding the adequacy of purely symptom-based categorical classification systems (Cuellar et al., 2005; Smith & Craddock, 2011).

Anxiety disorders collectively represent the most prevalent class of mental disorders. An estimated 301 million people worldwide were living with an anxiety disorder in 2019 (World Health Organization, 2023). Modelling suggests that approximately 71% of the global anxiety disorder burden could theoretically be averted if all affected individuals had access to optimal treatment (Santomauro et al., 2023).

Post-traumatic stress disorder (PTSD) affects an estimated 3.9% of the global population at some point in life, with substantially higher prevalence observed in conflict-affected, refugee, and military populations (Koenen et al., 2017). Complex post-traumatic stress disorder (C-PTSD), introduced in ICD-11, appears particularly prevalent among individuals exposed to prolonged or repeated interpersonal trauma. Recent meta-analytic estimates suggest pooled prevalence rates of approximately 4-6% in community samples, rising substantially in military, refugee, and clinical populations (Fung et al., 2025; Huynh et al., 2025). PTSD is also highly comorbid with depression, substance use disorders, chronic pain, and suicidality, contributing to elevated functional impairment and treatment complexity (Brady et al., 2000).

Attention-deficit/hyperactivity disorder (ADHD) has a worldwide pooled prevalence of approximately 5.3% in children and adolescents (Polanczyk et al., 2007), while adult ADHD prevalence is estimated at around 2.8% across high- and middle-income countries (Fayyad et al., 2017), with higher estimates when symptom-based rather than persistence-based criteria are applied (Song et al., 2021). ADHD is increasingly recognised as a lifespan condition associated with substantial functional impairment in adulthood and elevated rates of comorbid mood, anxiety, and substance use disorders (Faraone et al., 2021).

Bipolar spectrum disorders, including bipolar I disorder, bipolar II disorder, and cyclothymic disorder, have a lifetime prevalence of 2.4% worldwide (Merikangas et al., 2011). Bipolar disorders are associated with high levels of functional impairment, elevated suicide risk, substantial psychiatric and medical comorbidity, and frequent diagnostic delay or misclassification (Daveney et al., 2019; GBD 2019 Mental Disorders Collaborators, 2022). Depressive episodes account for the majority of symptomatic burden and longitudinal morbidity in bipolar disorder (Judd et al., 2002). Bipolar disorder is notable both for frequent under-detection and for substantial rates of overdiagnosis in clinical settings, particularly among individuals presenting with affective instability, trauma histories, personality pathology, or complex comorbidity (Daveney et al., 2019; Paris & Black, 2015).

Alcohol and substance use disorders are prevalent across all care settings and are associated with substantial medical, psychiatric, and social burden. Hazardous alcohol use affects a substantial proportion of adults globally, particularly in high-income countries, while drug use disorders affect an estimated 39.5 million people worldwide (United Nations Office on Drugs and Crime, 2024). Substance use disorders frequently co-occur with mood, anxiety, and psychotic disorders, contributing to greater illness severity, poorer treatment outcomes, and increased suicide risk (Hunt et al., 2020). Despite this, integrated screening and treatment for co-occurring psychiatric and substance use disorders remain inconsistently implemented across healthcare systems (Kelly & Daley, 2013).

Obsessive-compulsive disorder (OCD) has a lifetime prevalence of approximately 2–3% in the U.S. and

is associated with substantial functional impairment, reduced quality of life, and elevated rates of comorbid depression and anxiety (Ruscio et al., 2010). Delays between symptom onset and receipt of appropriate treatment frequently exceed a decade, in part because of stigma, concealment of symptoms, limited clinician recognition, and restricted access to exposure and response prevention therapy (Dell'Osso et al., 2013).

Psychotic-spectrum conditions, including schizophrenia and schizoaffective disorder, affect approximately 0.3% of the population worldwide (World Health Organization, 2025a). Schizophrenia is associated with substantially reduced life expectancy, with meta-analytic evidence suggesting a reduction of approximately 15-20 years relative to the general population (Peritogiannis et al., 2022). Assessment of psychotic-spectrum disorders typically requires approaches extending beyond self-report screening, including observational assessment and evaluation of cognitive and social cognitive functioning.

Feeding and eating disorders, including anorexia nervosa, bulimia nervosa, binge eating disorder, and other specified feeding or eating disorders (OSFED), are associated with among the highest mortality rates of any psychiatric disorders (Krug et al., 2025). OSFED is also among the most prevalent eating disorder diagnoses in community samples. Eating disorders are substantially under-detected in primary care and are highly comorbid with depression, anxiety, PTSD, and OCD.

Insomnia and sleep disorders are both independent conditions and transdiagnostic features that significantly influence the course and treatment response of nearly all other mental health conditions. Insomnia disorder alone affects an estimated 10–15% of the adult population, with substantially higher rates among individuals with mood, anxiety, and substance use disorders (Riemann et al., 2017).

Sexual dysfunction is highly prevalent across both sexes and is frequently comorbid with depression, anxiety, and the pharmacological treatments used to manage them. Female sexual dysfunction affects an estimated 40–50% of women at some point during their lives, and affects approximately 30% of men across the adult lifespan, with prevalence increasing with age (McCabe et al., 2016a, 2016b). Despite their prevalence and impact on quality of life, sexual dysfunctions are among the most under-screened conditions in both primary care and psychiatric settings.

Autism spectrum conditions have an estimated global median prevalence of approximately 1%, with growing recognition of underdiagnosis in females and adults (Zeidan et al., 2022). Autism is associated with elevated rates of comorbid anxiety, depression, ADHD, and eating disorders, and individuals on the autism spectrum frequently have unmet mental health needs due to assessment instruments and care pathways that were not designed for neurodivergent populations.

Subjective cognitive decline is relevant both as a potential early indicator of neurodegenerative conditions and as a clinical concern in its own right among older adults and individuals with mood disorders. Approximately 6.5 million Americans over 65 years of age have dementia (Alzheimer's Association, 2022). Brief screening for cognitive decline is increasingly recognised as an important component of comprehensive mental health assessment, particularly in primary care and residential care settings.

Personality disorders are characterised by enduring, inflexible patterns of inner experience and behaviour that deviate from cultural expectations, are pervasive across personal and social contexts, and cause significant distress or functional impairment. Borderline personality disorder (BPD) alone has an estimated prevalence of approximately 1–2% in the general population and substantially higher rates in clinical settings (Ellison et al., 2018; Leichsenring et al., 2024). Personality disorders are associated with elevated suicide risk and high rates of comorbidity with mood, anxiety, substance use, and eating disorders (Leichsenring et al., 2024). The relationship between categorical personality disorder diagnoses and dimensional personality traits remains actively debated, with the ICD-11 having moved to a primarily dimensional model based on severity and trait domains (Bach & First, 2018).

Adjustment disorders, stress-related presentations, and emotional dysregulation represent some of the most common presentations in primary care, employee assistance programme, and university counselling settings. Adjustment disorders are defined by a disproportionate response to an identifiable stressor, but in practice the boundary between adjustment disorder, subthreshold mood and anxiety symptoms, and normal stress responses remains poorly defined (O'Donnell et al., 2019). Chronic stress, burnout, and emotional dysregulation are clinically significant presentations that frequently drive help-seeking but may not meet criteria for a formal DSM-5 diagnosis. These presentations are important for a comprehensive assessment standard because they represent a large proportion of clinical encounters and because untreated subclinical distress is a recognised risk factor for progression to diagnosable disorders.

Body dysmorphic disorder (BDD) is classified within the obsessive-compulsive spectrum in DSM-5 and is characterised by preoccupation with perceived defects in physical appearance that are not observable or appear slight to others. BDD has an estimated point prevalence of approximately 2% in the general population, with substantially higher rates in cosmetic surgery and dermatology settings (Veale et al., 2016). It is frequently comorbid with depression, social anxiety, and eating disorders, and is associated with high levels of

functional impairment and elevated suicide risk relative to its prevalence (Angelakis et al., 2016).

Specific phobias and social anxiety disorder are among the most prevalent anxiety disorders, with cross-national lifetime prevalence estimates of approximately 7% for specific phobia (Wardenaar et al., 2017) and up to 12% for social anxiety disorder in high-income settings (Kessler et al., 2005b). Despite their prevalence, both are under-detected in routine clinical settings because patients often do not present with phobic symptoms as a primary complaint, and clinicians may not screen for them systematically.

Relationship difficulties and challenges in interpersonal functioning are among the most common reasons for seeking psychological help and are both a consequence of and a contributor to mental health conditions across diagnostic categories. Relationship discord has been shown to covary with a range of psychiatric disorders, to predict the incidence of mood, anxiety, and substance use disorders, and to be associated with poorer outcomes from individual-based treatments (Whisman, 2013). Loneliness and social isolation are similarly associated, bidirectionally, with depression and broader mental ill-health (Holt-Lunstad, 2024). Systemic and family therapy perspectives emphasise that mental health conditions do not occur in isolation from relational context: adjustment disorders are frequently triggered by relational stressors such as separation, bereavement, or family conflict; mood and anxiety disorders both cause and are exacerbated by relationship dysfunction; and child and adolescent mental health is inseparable from family system dynamics. A comprehensive assessment framework must therefore address interpersonal functioning as both a clinical domain in its own right and a transdiagnostic moderator of treatment outcomes across all other domains. While not a DSM-5 diagnostic category, interpersonal functioning is a critical treatment target and outcome domain that a reference standard cannot afford to omit.

In aggregate, GBD 2021 data indicate that over one billion individuals—approximately 14% of the global population—were living with at least one mental disorder in 2021, with mental disorders accounting for approximately one in six years lived with disability worldwide (World Health Organization, 2025b). These figures likely underestimate true prevalence, given well-documented under-detection, stigma-related underreporting, and data scarcity in low- and middle-income countries. The clinical scope of this standard reflects the reality that comprehensive mental health care cannot be delivered by addressing these conditions and presentations in isolation; effective assessment, treatment, and monitoring require technology-enabled systems capable of spanning the full range of diagnostic categories, symptom-level presentations, and functional domains that clinicians encounter in practice.

B Treatment Landscape

A standard concerned with mental health assessment must also address the clinical decisions that follow assessment: what treatment to offer, at what intensity, for which patient, and in what sequence. The mental health treatment landscape is broad and heterogeneous, shaped by differences in service delivery settings, evidence quality, treatment availability, cost, patient preference, side-effect burden, contraindications, and psychiatric and medical comorbidity. Evidence evaluation in psychiatry and psychotherapy is also affected by broader metascientific concerns, including publication bias, selective reporting, researcher allegiance effects, and financial or professional conflicts of interest (Ioannidis, 2008; Leichsenring et al., 2022).

Pharmacological treatment remains one of the most widely used interventions across psychiatric disorders. For depression and anxiety disorders, first-line pharmacological treatments commonly include selective serotonin reuptake inhibitors (SSRIs) and serotonin–norepinephrine reuptake inhibitors (SNRIs), with alternative antidepressant classes used in cases of non-response, intolerance, or specific clinical presentations (Cipriani et al., 2018; National Institute for Health and Care Excellence, 2022a). Meta-analytic evidence suggests that antidepressants are, on average, more effective than placebo, although effect sizes vary across individuals and clinical severity levels (Cipriani et al., 2018). Adverse effects and tolerability remain important considerations, including sexual dysfunction, weight gain, emotional blunting, discontinuation symptoms, metabolic effects, sedation, and dependency risks depending on medication class. Similar benefit–risk trade-offs apply across other psychiatric medications, including benzodiazepines, mood stabilisers, antipsychotics, stimulant medications, and medications for substance use disorders.

Psychotherapy encompasses a diverse set of structured interventions, including cognitive behavioural therapy (CBT), dialectical behaviour therapy (DBT), trauma-focused therapies, acceptance and commitment therapy (ACT), mindfulness-based interventions, interpersonal psychotherapy, psychodynamic therapies, and family or systemic approaches. CBT and related behavioural interventions are among the most extensively studied and widely recommended psychotherapies across mood, anxiety, trauma-related, and other common psychiatric disorders (American Psychological Association, 2019; Cuijpers et al., 2020; National Institute for Health and Care Excellence, 2022a).

Meta-analytic evidence supports the efficacy of multiple psychotherapeutic modalities across psychiatric conditions, although important methodological limitations in the evidence base have also been documented. Effect sizes are often larger in studies using waiting-list or minimal-treatment control groups than in studies

using active comparators, and researcher allegiance effects may contribute to outcome differences across trials (Cuijpers et al., 2016; Leichsenring et al., 2022; Munder et al., 2013). Comparative studies between established psychotherapies frequently report relatively small differences in efficacy, while combined psychotherapy and pharmacotherapy is often associated with greater benefit than either treatment alone for some disorders and severity levels (Cuijpers et al., 2019; Leichsenring et al., 2022). Access to psychotherapy also remains uneven across health systems because of workforce shortages, financial barriers, long waiting times, and geographic disparities in service availability.

Non-pharmacological and neuromodulatory interventions targeting neural systems include electroconvulsive therapy (ECT), repetitive transcranial magnetic stimulation (rTMS), ketamine and esketamine, and emerging psychedelic-assisted therapies. These interventions are most commonly considered in treatment-resistant or severe conditions, particularly major depressive disorder (Carhart-Harris & Goodwin, 2017; Malhi & Mann, 2018). ECT remains among the most effective acute interventions for severe depression, particularly in psychotic depression and severe suicidality. Evidence also supports the efficacy of rTMS and ketamine-based interventions in treatment-resistant depression, though important questions remain regarding durability of response, long-term safety, implementation costs, and scalability within routine clinical systems (Bahji et al., 2021; Malhi & Mann, 2018).

Digital therapeutics deliver structured psychological interventions through software platforms, commonly based on cognitive behavioural therapy (CBT), behavioural activation, or related behavioural frameworks. Internet-delivered and app-based interventions have demonstrated efficacy for common mental health conditions, particularly depression and anxiety disorders, in controlled research settings (Carlbring et al., 2018; Wright & Caudill, 2020). Several digital therapeutics have also received regulatory clearance for specific indications. Their potential advantages include scalability, lower marginal delivery costs, and improved accessibility for underserved populations. However, real-world implementation studies suggest that treatment engagement, adherence, and sustained retention are substantial challenges, and effectiveness outside controlled trial settings is often lower than efficacy observed in research environments (Baumel et al., 2019; Torous et al., 2021).

Lifestyle and adjunctive interventions, including exercise, sleep interventions, dietary modification, and social or behavioural activation approaches, are increasingly incorporated into prevention and treatment strategies for psychiatric disorders. Meta-analytic evidence supports modest beneficial effects of interventions such as structured exercise and sleep-focused treatments for some conditions, particularly depression and anxiety disorders (Singh et al., 2023). Social prescribing and peer-support approaches have also been adopted in several health systems as part of broader community-based mental healthcare models, although the evidence base remains heterogeneous and methodologically variable (Bickerdike et al., 2017).

Across treatment modalities, most interventions in psychiatry demonstrate moderate average effect sizes, substantial between-patient variability in response, and evidence bases affected by methodological limitations and risk of bias (Ioannidis, 2008; Leichsenring et al., 2022). These limitations have direct implications for technology-supported clinical decision-making systems. Clinical guidelines and evidence summaries cannot be treated as deterministic rules applicable uniformly across patients. Instead, decision-support systems should represent uncertainty explicitly, incorporate evidence quality and provenance, and support shared decision-making that integrates efficacy, adverse effects, contraindications, availability, cost, and patient preferences. The appropriate role of such systems is to support structured and evidence-informed clinical reasoning rather than automate treatment decisions.

C Measurement Modalities

The conditions described above can be assessed through a range of measurement modalities, each with distinct psychometric properties, feasibility constraints, and evidence bases. A reference standard for technology-powered mental health care must accommodate this diversity, since no single modality is optimal across all conditions, settings, and populations. The following taxonomy summarises the principal approaches; we then provide detailed assessment of the evidence base for each, and specify system requirements for their integration.

Paper-and-pencil self-report instruments have historically been the dominant modality for patient-reported outcome and mental health symptom assessment, and many widely used psychiatric instruments were originally developed and validated in paper format (Coons et al., 2009; Gwaltney et al., 2008). Paper administration requires minimal technological infrastructure and remains important in low-resource or digitally constrained settings. However, paper-based workflows create practical limitations for scalable longitudinal assessment: scoring and data entry require staff time, introduce opportunities for error, and make real-time feedback, automated routing, remote completion, and integration with electronic health records more difficult.

Computer- and tablet-administered self-report questionnaires translate self-report assessment into digital workflows, enabling automated scoring, conditional branching, longitudinal tracking, remote completion, and integration with clinical information systems.

Systematic reviews generally support equivalence between paper and electronic administration for many patient-reported outcome measures, provided that migration preserves item content, response options, layout, and administration conditions; however, equivalence should not be assumed when substantial format changes are introduced or when instruments are used in populations not included in validation studies (Coons et al., 2009; Gwaltney et al., 2008; Muehlhausen et al., 2015; O'Donohoe et al., 2023).

Mobile self-report extends digital assessment to smartphones and other personal devices, allowing patients to complete measures outside clinical encounters and supporting repeated assessment, reminders, and measurement-based care workflows. These features are particularly relevant for longitudinal monitoring, stepped care, and symptom feedback in routine mental healthcare (Fortney et al., 2017). However, mobile implementation also introduces design and equity considerations, including usability on small screens, device access, digital literacy, accessibility needs, privacy, and the need to preserve measurement comparability when migrating instruments across modes (Coons et al., 2009; O'Donohoe et al., 2023). Formal guidance exists for electronic migration and implementation of patient-reported outcome measures, and these standards should be followed when adapting paper instruments for digital use.

Clinician-administered structured and semi-structured interviews (e.g., the Structured Clinical Interview for DSM Disorders [SCID] and the Mini International Neuropsychiatric Interview [MINI]) remain widely used reference methods for diagnostic assessment and psychiatric research (First et al., 2016; Sheehan et al., 1998). These formats allow clinicians to clarify ambiguous responses, evaluate symptom context, assess behavioural presentation, and integrate clinical judgment in ways not possible with fixed self-report measures alone. Interview-based assessment is particularly important for complex differential diagnosis, high-risk presentations, and populations in whom self-report may be unreliable or insufficient. However, structured interviews are resource-intensive, requiring clinician time, training, and ongoing attention to inter-rater reliability (Regier et al., 2013). Technology-supported systems may improve workflow efficiency through structured prompts, automated scoring, and documentation support while preserving clinician oversight and judgment.

Computerised cognitive and neuropsychological tasks assess domains such as attention, processing speed, executive function, working memory, and cognitive flexibility that are relevant across psychiatric, neurodevelopmental, and neurodegenerative disorders (Bilder & Reise, 2019). Common examples include computerised adaptations of established neuropsychological paradigms such as the Wisconsin Card Sorting Test, Trail Making Test, continuous performance tasks, and standardised emotion-recognition paradigms. These approaches generate performance-based measures including accuracy, reaction time, and error patterns that complement symptom self-report and clinical interview data. Digital administration can improve standardisation, automate scoring, and enable precise temporal measurement, although valid interpretation still depends on appropriate testing conditions, device characteristics, normative reference data, and consideration of practice effects in repeated assessment (Bilder & Reise, 2019; Cernich et al., 2007). Several major neuropsychological assessment batteries, including digital versions of Wechsler cognitive and memory assessments, have also transitioned toward digitally assisted or fully digital administration.

Gamified and serious-game assessments embed cognitive or behavioural measurement within interactive task environments intended to improve engagement and support repeated assessment. These approaches have been explored in areas including ADHD assessment, paediatric neuropsychology, cognitive screening, and mental health monitoring (Lumsden et al., 2016). Early evidence suggests that gamified approaches may improve user engagement and tolerability relative to conventional testing formats, although the psychometric evidence base remains substantially less mature than that for traditional neuropsychological instruments and many systems still require further validation against established clinical measures (Lumsden et al., 2016).

Speech and voice analysis uses acoustic and linguistic features of speech – including prosody, pitch variability, speech rate, pause structure, articulation, and lexical characteristics – as potential markers of mental health and neurological conditions. Research has identified associations between speech features and disorders including depression, bipolar disorder, psychotic-spectrum disorders, anxiety disorders, and cognitive decline (Cummins et al., 2015; Low et al., 2020). Speech data may be collected through structured speech tasks or extracted from conversational audio recorded during clinical encounters, telehealth sessions, or mobile interactions. The modality is attractive because it is relatively low burden, non-invasive, and compatible with remote and longitudinal monitoring workflows. However, despite substantial research interest, evidence for reliable clinical deployment remains limited, and important challenges remain regarding generalisability across languages, dialects, recording environments, devices, demographic groups, and clinical contexts (Low et al., 2020; Torous et al., 2021).

Eye tracking and gaze analysis measure patterns of visual attention and ocular behaviour, including fixation duration, saccadic movements, smooth pursuit, blink rate, and pupil dilation, which may reflect cognitive, attentional, emotional, and social processing. Eye-

tracking research has identified group-level differences associated with autism spectrum conditions, ADHD, anxiety disorders, depression, and psychotic-spectrum disorders (Shic, 2016). Historically, eye tracking required specialised laboratory hardware, but consumer-grade eye tracking and camera-based gaze estimation are increasingly available on mobile devices, laptops, and virtual reality systems. Although these developments improve accessibility and scalability, lower-cost hardware and unconstrained environments introduce additional sources of measurement noise, calibration variability, and reduced precision that may affect reliability and clinical validity (Holmqvist, Nyström, et al., 2023).

Facial expression and video-based analysis uses computer vision to quantify facial behaviour, affective display, head movement, posture, and other observable nonverbal signals from video collected during structured tasks, clinical interviews, or telehealth interactions. This modality overlaps with emotion-recognition assessment, where patients identify affective cues in others, but is distinct from passive analysis of the patient's own expressive behaviour. Video-based features have been studied in relation to depression, psychotic-spectrum disorders, neurodevelopmental conditions, and cognitive impairment, and can be combined with speech or physiological data in multimodal assessment pipelines (Cohn et al., 2018; Jiang et al., 2022). However, current evidence is strongest in research settings, and clinical deployment remains limited by small samples, heterogeneous tasks, dataset bias, privacy requirements, consent, and uncertainty about generalisation across demographic groups, recording environments, and care settings (Martinez-Martin et al., 2018, 2021).

Multimodal sensor fusion combines two or more data streams – for example speech, facial behaviour, gaze, movement, keyboard interaction, smartphone use, or physiological signals – to support assessment or monitoring. Multimodal approaches are motivated by the fact that mental health presentations are heterogeneous and expressed across multiple behavioural, cognitive, affective, and physiological channels. Reviews of digital phenotyping and multimodal behavioural assessment suggest that combining complementary signals may improve prediction or monitoring in some contexts, but the evidence base remains methodologically heterogeneous and is not yet sufficient to support broad diagnostic use (Cohn et al., 2018; Moura et al., 2023; Torous et al., 2021). Translation from research-grade systems to routine clinical or consumer-device settings requires validation across devices, populations, languages, environments, and clinical workflows, as well as clear data governance and evidence standards.

Digital phenotyping and passive smartphone sensing collects behavioural and physiological data passively from smartphones and connected devices without requiring active user engagement. Data streams include GPS-derived mobility patterns, accelerometer-based activity and sleep estimates, screen usage patterns, communication metadata (call and text frequency and timing), and app usage. Research has demonstrated associations between passive digital phenotypes and symptom trajectories across depression, anxiety, bipolar disorder, and psychotic-spectrum disorders, with emerging evidence suggesting that smartphone-derived behavioural signals may help predict symptom exacerbations before they become clinically apparent (Jacobson & Bhattacharya, 2022; Onnela & Rauch, 2016). However, the field faces significant challenges: variability in sensor quality across devices, lack of standardisation in feature extraction, limited replication of findings across studies, and substantial privacy concerns regarding continuous behavioural monitoring. The evidence base currently supports digital phenotyping as a research tool with clinical promise rather than a validated clinical assessment modality.

Ecological momentary assessment (EMA) and experience sampling use repeated brief assessments delivered in daily life, typically via smartphone, to capture symptoms, mood, behaviour, and context close to the time they occur. EMA was developed to address limitations of retrospective self-report, including recall bias and the difficulty of reconstructing temporal and contextual variation from global symptom summaries (Shiffman et al., 2008). In mental health research, EMA and experience sampling methods are used to characterise within-person variability, affective dynamics, stress reactivity, diurnal patterns, and context-dependent symptom fluctuations (Myin-Germeys et al., 2018; Stone et al., 2023). EMA can be implemented using time-based, event-based, or hybrid sampling schedules, and may be combined with passive sensing or wearable physiological data. Its clinical promise lies in supporting longitudinal monitoring and personalised case formulation, but implementation requires attention to participant burden, adherence, privacy, missing data, and the translation of intensive longitudinal data into clinically interpretable outputs.

Wearable physiological monitoring via smartwatches, fitness trackers, and research-grade wearable devices can capture signals such as movement, sleep–wake patterns, heart rate, heart-rate variability, electrodermal activity, and skin temperature. In mental health research, these data streams are commonly studied as components of digital phenotyping and as potential markers of sleep disturbance, activity change, circadian disruption, stress physiology, and symptom fluctuation (Aledavood et al., 2019; Bufano et al., 2023; Torous et al., 2021). Consumer-grade wearables offer scalable and relatively low-burden data collection in naturalistic settings, but clinical interpretation remains limited by device variability, measurement error, missing data, pro-

prietary algorithms, population differences, and the lack of validated normative frameworks for most psychiatric use cases (Aledavood et al., 2019; Bufano et al., 2023).

Continuous monitoring in clinical environments such as inpatient psychiatric units, residential treatment facilities, and intensive outpatient programmes can combine structured assessments, staff observations, wearable devices, environmental sensors, and electronic health record data. These settings are potentially data-rich, but the evidence base for automated real-time risk detection in psychiatry remains early. Current and proposed applications include supporting observation workflows, identifying behavioural change over time, improving documentation, and informing risk review, rather than replacing clinical judgment. Implementation must account for integration with nursing workflows, false alarms and alarm fatigue, privacy, informed consent, data governance, and the ethical implications of monitoring in settings where patient autonomy may already be constrained (Martinez-Martin et al., 2018; Torous et al., 2021).

Structural and functional neuroimaging (MRI, fMRI, PET, and SPECT) can provide information about brain structure, function, connectivity, and neurochemistry. In clinical psychiatry, structural neuroimaging is most often used when clinically indicated to evaluate possible neurological or medical contributors to psychiatric presentations, and it is central to the assessment of many neurodegenerative conditions. Functional neuroimaging has generated important research findings across depression, schizophrenia, PTSD, OCD, autism, ADHD, and related conditions, and is widely studied for treatment-response prediction and biomarker discovery. However, neuroimaging remains expensive, requires specialised facilities, and has limited utility as a standalone diagnostic tool for most psychiatric disorders. Reviews and consensus reports emphasise that psychiatric neuroimaging biomarkers are promising but generally not yet sufficiently validated for routine individual-level diagnosis or treatment selection (Abi-Dargham et al., 2023; First et al., 2012; Schnack & Kahn, 2016). Reproducibility and generalisability concerns, including small sample sizes and analytic variability in fMRI research, further support treating neuroimaging as a specialist adjunctive modality rather than a scalable screening or routine monitoring tool (Botvinik-Nezer et al., 2020; Turner et al., 2018).

Electroencephalography (EEG) and neurophysiological assessment measure electrical brain activity with high temporal resolution. EEG is an established clinical tool in the evaluation of seizure disorders and altered states of consciousness, and standardised guidance exists for routine and sleep EEG recording and reporting (Peltola et al., 2023; Tatum et al., 2016). In psychiatry, quantitative EEG (qEEG) and event-related potentials (ERPs) have been studied as candidate biomarkers in conditions such as ADHD, depression, psychosis, and traumatic brain injury, including for treatment-response prediction. However, these applications remain less clinically established than EEG's neurological uses, and several proposed psychiatric EEG markers have shown variable reproducibility or limited diagnostic specificity (Gloss et al., 2016; Simmatis et al., 2023; Widge et al., 2019). Consumer EEG devices and portable systems may expand access to neurophysiological measurement, but reduced channel count, signal quality, artefact susceptibility, and limited clinical validation mean that EEG is best treated here as a specialist adjunctive modality rather than a standalone psychiatric assessment tool.

Natural language processing (NLP) of clinical and patient-generated text analyses unstructured language from clinical notes, assessment narratives, therapy transcripts, patient journals, digital messages, and social media. NLP methods range from lexicon-based approaches to machine-learning and large-language-model systems, and have been studied for symptom extraction, mental health classification, risk detection, suicide-risk modelling, and analysis of psychotherapy or other mental health interactions (Dreisbach et al., 2019; Malgaroli et al., 2023; Zhang et al., 2022). NLP may complement structured assessment by identifying clinically relevant information expressed in free text but not captured in fixed-response instruments. However, clinical deployment remains constrained by dataset bias, transportability across settings, privacy and consent requirements, interpretability, and the need for prospective validation in real-world workflows (Spasic & Nenadic, 2020; Torous et al., 2021) data ownership, and the risk of over-interpreting linguistic patterns.

These modalities exist on a spectrum from well-validated to experimental, and from episodic to continuous. The current evidence base is strongest for self-report instruments (paper and digital) and clinician-administered interviews, which form the foundation of the requirements described in this document. Computerised cognitive tasks occupy a middle ground with established research utility and growing clinical application. The technology-intensive modalities - speech analysis, eye tracking, digital phenotyping, multimodal sensor fusion, and NLP - represent the frontier of mental health measurement, with significant research momentum but evidence that has not yet matured to the level required for standalone clinical deployment. A reference standard should be designed to accommodate the full range of modalities as the evidence base evolves, with clear evidence-tier requirements governing which modalities are appropriate for which clinical functions and at what stage of validation. The choice of modality in any given deployment is also fundamentally shaped by the care setting: what is feasible in a research laboratory with calibrated hardware differs from what is

achievable on a patient's personal smartphone, in a primary care waiting room, via a wearable during daily life, or in a residential treatment facility with continuous monitoring infrastructure.

D Care Settings

The challenges outlined above are not confined to a single clinical context. They appear across the settings in which mental health care is sought, delivered, or needed but unavailable. The following summary illustrates domains in which technology-augmented assessment, decision support, and care navigation may be relevant.

Primary care is a major point of identification and treatment for common mental health conditions in the United States, particularly where access to specialty mental healthcare is limited. Primary care workflows are time constrained and must address mental health alongside multiple competing medical priorities. Depression screening in adults is recommended when systems are in place for accurate diagnosis, effective treatment, and appropriate follow-up (US Preventive Services Task Force, 2023). In practice, pathways from screening to diagnostic assessment, treatment initiation, monitoring, or referral vary across clinics and health systems. Collaborative-care models provide one of the strongest evidence bases for integrating mental health treatment into primary care, combining systematic identification, care management, psychiatric consultation, and measurement-based follow-up (Fortney et al., 2017; Unützer et al., 2002). For technology-supported assessment to be useful in primary care, it should minimise added clinician burden, support pre-visit or point-of-care screening, automate scoring, integrate results into clinical workflows, and produce decision support aligned with primary-care scope and referral pathways.

Emergency departments and acute care settings manage high-acuity mental health presentations under substantial time pressure, including suicidality, self-harm, intoxication, agitation, psychosis, and acute safety concerns. Emergency workflows prioritise immediate safety assessment, medical stabilisation, risk stratification, disposition planning, and handoff to inpatient, crisis, or outpatient services. Psychiatric boarding in emergency departments is a persistent systems problem in the United States, reflecting limited availability of appropriate psychiatric inpatient and crisis-care capacity (Chang et al., 2012; Nordstrom et al., 2019). Evidence from a large US emergency-department suicide-prevention study suggests that structured screening, safety planning, and follow-up interventions can improve identification and reduce subsequent suicidal behaviour in some populations (Miller et al., 2017). In this context, technology-supported assessment should be brief, safety-focused, integrated with triage and EHR workflows, and designed to improve documentation and continuity of information across discharge or transfer. Comprehensive multi-domain assessment is usually less appropriate in the ED than targeted risk screening, structured documentation, and reliable handoff.

Outpatient psychiatry and specialty mental health settings provide assessment and treatment for patients with more complex, severe, recurrent, or treatment-resistant mental health conditions. These settings are well suited to longitudinal assessment because patients often receive repeated medication-management, psychotherapy, or combined-treatment visits over time. Measurement-based care in psychiatry is defined by the repeated use of validated symptom, functioning, and safety measures to inform clinical decisions, and has been recommended as a way to improve treatment monitoring and outcomes (American Psychiatric Association, 2023; Fortney et al., 2017). However, implementation remains uneven across routine mental health settings, with barriers including workflow burden, measure selection, clinician engagement, EHR integration, and the need to translate scores into actionable treatment decisions (Fortney et al., 2017; Lewis et al., 2019). For technology-supported assessment, outpatient psychiatry therefore represents a natural use case for pre-visit digital assessment, automated severity tracking, treatment-response monitoring, side-effect monitoring, and decision support that remains under clinician control.

Older adults and residential or long-term care populations require assessment approaches that account for multimorbidity, polypharmacy, cognitive impairment, sensory impairment, functional limitations, and reliance on carers or staff for collateral information. Depression and cognitive impairment may be difficult to detect in long-term care settings, particularly when symptoms overlap with dementia, medical illness, apathy, delirium, or medication effects (Brown et al., 2009; Volicer et al., 2011). Medication review is also central in geriatric care because potentially inappropriate medications and drug–disease interactions are common safety concerns in older adults. Cognitive assessment in older adults should be interpreted as part of a broader clinical evaluation rather than as a standalone diagnosis (National Institute on Aging, 2023; US Preventive Services Task Force, 2020). Digital self-report may be useful for some older adults, but systems should support alternative administration modes, including clinician-assisted completion, informant-report instruments, accessible interfaces, and integration with cognitive and functional assessment workflows.

Telehealth and asynchronous care expanded rapidly during the COVID-19 pandemic and has remained especially important in mental health care in the United States. Recent US studies show that telehealth availability for mental health services remains substantially above pre-pandemic levels, although availability varies across facilities and service types (Cantor et al., 2024; McBain et al., 2024). Telepsychiatry

and telepsychotherapy can support screening, follow-up assessment, medication management, psychotherapy delivery, and longitudinal monitoring when clinically appropriate. Meta-analytic evidence suggests that live video psychotherapy can produce outcomes comparable to in-person psychotherapy for several common conditions, particularly when structured evidence-based treatments are delivered with adequate fidelity (Fernandez et al., 2021). Important limitations remain for interventions requiring in-person delivery or monitoring, such as ECT, TMS, ketamine or esketamine administration, inpatient stabilisation, some physical examination components of medication monitoring, and situations requiring urgent in-person risk management. In the United States, prescribing controlled substances by telemedicine remains governed by federal and state rules that have continued to evolve after the COVID-19 public health emergency (Drug Enforcement Administration & Department of Health and Human Services, 2025). Remote and asynchronous modalities also require attention to digital access, privacy, clinical appropriateness, acute-risk workflows, and variation in platform quality.

Employee assistance programmes (EAPs) and workplace mental health services are employer-sponsored benefits that commonly provide confidential, short-term counselling, supportive resources, and referral to external care when indicated. In the United States, EAPs have traditionally been offered as standalone or externally delivered benefits with a limited or fixed number of free services, although programme structure varies across employers and vendors. Evidence suggests that EAP counselling may improve psychological distress, work functioning, absenteeism, and presenteeism among employees who use these services, but the evidence base is limited by observational designs, heterogeneous programme models, and variable outcome measurement (Attridge & Dickens, 2022; Bouzikos et al., 2022). For technology-supported assessment, EAPs may function as a high-volume, low-barrier entry point for brief screening, triage, referral navigation, and outcome monitoring. However, systems must preserve confidentiality boundaries between employees and employers, avoid inappropriate disclosure of individual clinical information, and support referral continuity when an employee requires care beyond the EAP scope.

University and college counselling centres face rapidly growing demand, with prevalence of anxiety, depression, and suicidality among students rising significantly over the past decade (Association, 2023). Many counselling centres operate under substantial resource constraints relative to student demand, contributing to wait times and limits on service availability (Association for University and College Counseling Center Directors, 2023).

Forensic, correctional, and justice-involved populations have high rates of serious mental illness and complex mental health needs, while correctional mental health services vary substantially in screening, triage, assessment, intervention, and re-entry planning (Fazel & Seewald, 2012; Simpson et al., 2022). Continuity of mental health care following release is often poor, particularly for people with recurrent or brief periods of imprisonment (Browne et al., 2022).

Wellness, consumer mental health, and self-managed care represents a large and growing population engaging with mental health support outside formal health systems. This includes consumer mood-tracking apps, meditation and mindfulness platforms, AI-powered conversational agents, and digital self-help programmes based on CBT and related therapeutic frameworks (Torous et al., 2020). These tools may support access, monitoring, and self-management, but they also raise quality, evidence, privacy, and governance concerns, particularly because many consumer-facing tools are not evaluated to the same standards as clinical interventions (Larsen et al., 2019).

Veterans, military, and high-complexity populations often present with complex and overlapping mental health needs, including PTSD, traumatic brain injury, substance use, and transition-related barriers to care. Mild traumatic brain injury among deployed soldiers has been strongly associated with PTSD and other post-deployment health problems, and recent review evidence identifies multiple systemic, logistical, and social barriers to mental health service access during the transition from military to civilian life (Ein et al., 2024; Hoge et al., 2008).

Children, adolescents, and neurodevelopmental populations require developmentally appropriate assessment pathways that often combine reports from multiple informants, including children, parents or guardians, and teachers. Multi-informant assessment is a core feature of child and adolescent mental health evaluation because symptoms may vary across contexts such as home and school (De Los Reyes et al., 2015). For children and adolescents with neurodevelopmental conditions, assessment should not rely solely on standard digital self-report; alternative pathways may include parent- or teacher-report measures, clinician-administered assessment, observational information, and adapted interfaces.

Substance use and comorbidity pathways illustrate the consequences of siloed care models. Co-occurring mental health and substance use disorders are common, and integrated treatment models emphasize screening and treatment planning across both domains rather than requiring people to move between separate mental health and substance-use programmes (Substance Abuse and Mental Health Services Administration, 2010).

Payers and utilisation management increasingly require measurement-based care as a condition of re-

imbursement, yet the instruments and reporting infrastructure needed to support this requirement are not consistently available across providers, particularly in behavioural health where outcome measurement has historically lagged behind physical medicine (Knutson et al., 2021). Value-based payment models depend on standardised outcome data that most behavioural health systems do not routinely collect.

Global and low-resource settings face the most acute workforce shortages: the median number of mental health workers per 100,000 population is roughly 1.4 in low-income countries and 3.8 in lower-middle-income countries, compared with over 60 in high-income countries (World Health Organization, 2021b). Task-shifting to non-specialist and community health workers has been the primary strategy for expanding access in these contexts and is the core framework underpinning the WHO mhGAP programme (Patel et al., 2018). Digital platforms have particular potential here but require cultural adaptation, offline capability, and instruments validated in local populations (Heim et al., 2021).

Language- and culturally diverse populations are underserved across all of the above settings. Cultural explanatory models of illness, locally salient idioms of distress, instrument translation validity, and interpreter availability intersect to produce systematic differences in how symptoms are expressed, detected, and treated (Bhui & Bhugra, 2002).

The Collaborative Care Model (CoCM) is one of the most extensively evaluated frameworks for integrating behavioural health into primary care. CoCM structures the collaboration between primary care providers, behavioural health care managers, and consulting psychiatrists around measurement-based care: systematic screening, registry-based tracking, and treat-to-target protocols with defined escalation for non-response. A Cochrane review of 79 randomised controlled trials involving over 24,000 participants found that collaborative care produced greater improvements in depression and anxiety outcomes than usual care, with benefits sustained up to two years (Archer et al., 2012). However, CoCM adoption remains limited by the operational infrastructure it requires – registries, standardised instruments, care manager workflows, and psychiatric consultation capacity – that many health systems lack. Technology-powered systems that provide structured assessment, longitudinal monitoring, treatment response tracking, and clinician alerting directly support the core operational requirements of CoCM and similar stepped and integrated care models, potentially lowering the implementation barrier and enabling wider adoption.

The common thread across these settings and care delivery models is that each would benefit from the same core infrastructure: structured, validated, multi-domain assessment; longitudinal monitoring; evidence-based decision support; and systematic care navigation. The requirements defined in this document are designed to be adaptable across these contexts, with setting-specific configuration rather than setting-specific technology.

E Financial Structures and Economic Considerations

A reference standard that cannot sustain itself financially will not be adopted regardless of its clinical merit. This section addresses the economic dimensions of technology-powered mental health infrastructure: the costs of implementation and operation, the reimbursement mechanisms that can support them, and the economic case for investment at the health system, payer, and societal levels. A structural challenge is that the entities bearing implementation costs are not always those that realise downstream savings, creating misaligned incentives that can slow adoption despite system-level benefit (Knapp & Wong, 2020).

The economic burden of inadequate mental health infrastructure. Mental health conditions impose substantial direct and indirect economic costs, with global estimates placing the value of losses attributable to mental disorders in the trillions of US dollars, driven predominantly by lost productivity rather than direct treatment costs (Arias et al., 2022; Knapp & Wong, 2020). Direct costs include emergency presentations and inpatient admissions that may be preventable through earlier detection and management, repeated diagnostic workups due to misdiagnosis, and treatment of conditions that have progressed beyond early intervention windows. Indirect costs include absenteeism, presenteeism, disability claims, and the broader societal burden of untreated illness. The economic case for structured assessment infrastructure rests on the premise that systematic screening, longitudinal monitoring, and evidence-based care navigation can reduce these downstream costs by enabling earlier detection, accelerating treatment response through measurement-based adjustment, and preventing escalation to high-acuity care settings.

Reimbursement pathways for measurement and monitoring. The financial viability of different assessment modalities varies based on existing reimbursement structures. In the US, Current Procedural Terminology code 96127 reimburses brief emotional or behavioural assessment using validated standardised instruments (e.g., PHQ-9, GAD-7), with scoring and documentation per instrument, providing a direct billing pathway for digital self-report screening (American Medical Association, 2024a). The Collaborative Care Model (CoCM) codes 99492 – 99494 explicitly reimburse measurement-based care workflows including screening, registry tracking, and psychiatric consultation (American Medical Association, 2024b). These codes represent the most direct reimbursement pathway for the infrastructure described in this standard. In

contrast, emerging modalities such as speech analysis, eye tracking, digital phenotyping, and consumer wearable monitoring lack established reimbursement pathways in most jurisdictions, and regulatory and financial frameworks remain a recognised barrier to their clinical adoption (Torous et al., 2025). Clinician-administered structured interviews are reimbursable as part of clinical encounters but carry the full cost of clinician time. The reimbursement landscape therefore favours the digital self-report and measurement-based care approach that forms the foundation of this standard.

Cost structures by modality. Implementation and per-administration costs differ substantially across assessment modalities. Digital self-report instruments range from freely available (e.g., PHQ-9, GAD-7, AUDIT) to modest per-use licensing fees, with infrastructure costs limited to delivery, scoring, and EHR integration. Ecological momentary assessment shares this cost profile when delivered via smartphone. Clinician-administered interviews carry the full cost of clinician time, making them the most expensive per-administration modality despite requiring no specialised equipment. Consumer-grade wearables and smartphone-based passive sensing have been highlighted as more accessible alternatives to research-grade devices for digital phenotyping in mental health, though evidence on clinical utility remains preliminary (Hassan et al., 2025). We also note that many smartphone health applications (e.g., Apple Health, Google Health Connect) collect passive data without the need for additional wearable devices. Speech analysis and eye tracking require either specialised hardware or validated software running on consumer devices, with per-administration costs that are currently difficult to estimate due to the absence of standardised commercial offerings. Neuroimaging remains the most expensive modality by a wide margin. These cost structures reinforce the convergence argument: the modality that is both most clinically ready and least expensive per administration is digital self-report, which can be delivered at marginal cost approaching zero.

Value-based care and the economic case for measurement infrastructure. The transition from fee-for-service to value-based payment models creates a direct financial incentive for the infrastructure described in this standard, as providers under value-based contracts must demonstrate outcomes using validated measures rather than service volume (Knutson et al., 2021). Structured screening that detects conditions earlier, measurement-based care that identifies treatment non-response sooner, and care navigation that reduces dropout all support the metrics on which value-based contracts are evaluated; meta-analytic evidence indicates that measurement-based care is associated with greater symptom reduction and higher response rates compared to standard care for depression (Zhu et al., 2021). For payers, standardised outcome data provides the measurement infrastructure needed to design and adjudicate value-based contracts. The economic case for health systems is that the cost of implementing digital screening and monitoring infrastructure is substantially lower than the cost of the downstream failures it prevents; long-term follow-up of the IMPACT collaborative care trial found that the intervention was associated with lower total healthcare costs over four years compared with usual care, with the cost offset extending well beyond the active intervention period (Unützer et al., 2008).

Financial sustainability and self-supporting infrastructure. For the standard to achieve widespread adoption, the infrastructure it describes must be financially self-supporting rather than dependent on grant or philanthropic subsidy. Three mechanisms support this: direct reimbursement for screening and MBC activities via existing CPT and CoCM codes; cost avoidance through reduced acute utilisation and faster treatment response, as demonstrated in collaborative care trials (Unützer et al., 2008); and the value of the real-world evidence generated by the infrastructure for research, quality improvement, and payer reporting. In low- and middle-income settings, the economic argument shifts toward task-shifting – redistributing assessment and monitoring tasks from scarce specialist clinicians to digital infrastructure or trained non-specialist personnel – with WHO-led modelling estimating benefit-to-cost ratios in the range of 2.3 – 5.7 to 1 for scaled-up treatment of depression and anxiety once both economic productivity and health returns are considered (Chisholm et al., 2016).