Viewpoint

Measure the Ward First: Intentional Sufficiency as a Design Rule for Sleep Sensing in Acute Inpatient Psychiatry

Hong-Ming Chen[1,2], MD; Terry B-J Kuo[2], PhD, MD

[1]Department of Psychiatry, Chiayi Chang Gung Memorial Hospital, Puzi City, Chiayi County, Taiwan

[2]Institute of Brain Science, National Yang Ming Chiao Tung University, Taipei, Taiwan

Corresponding Author:
Terry B-J Kuo, PhD, MD
Institute of Brain Science
National Yang Ming Chiao Tung University
No. 155, Sec. 2, Linong St.
8F, Library and Information Building
Taipei, 112304
Taiwan
Phone: 886 2 2826 7058
Email: tbjkuo@nycu.edu.tw

Abstract

This Viewpoint argues that measurement design, not model capability, is the first constraint on clinical AI in acute inpatient psychiatry and proposes intentional sufficiency as a design rule for sleep sensing on acute wards. Acute psychiatric wards exclude much of the conventional sensing design space: ligature, ingestion, and projectile hazards; unreliable adherence; impaired consent; and nursing workload mean that a device that is excellent when a patient is cooperative may be absent when the patient is most unwell. We define intentional sufficiency in 3 parts: the intended clinical use sets the required signal properties and acceptable error; safety, validity, consent, and workflow constraints define the feasible design space; and within that space, the rule selects the least burdensome option that preserves the required properties, because excess capability can be paid for in persistence, not only in money. We contrast this rule with capability-maximizing design using 2 examples at opposite ends of the design space: a structured nightly nursing sleep duration record that, in Taiwan, already exists within routine care, and a disposable wrist-worn sensing architecture under development that reduces the wearable to what the derived sleep measure requires. The first shows when the rule can conclude that no hardware should be added, provided adequacy has been shown for the intended use; the second shows when it concludes that hardware must be redesigned, and is presented as a design prediction rather than a result. Importantly, agreement with a reference method does not by itself establish task-level sufficiency; greater sensing capability does not necessarily produce more clinically available information when deployment and persistence are constrained; and some signal properties are validity defining and cannot be traded against cost.

(J Med Internet Res 2026;28:e112674) doi: 10.2196/112674

KEYWORDS

inpatient psychiatry; measurement infrastructure; wearable sensors; artificial intelligence; AI; validation

The Measurement Problem Before the AI Problem

An AI system expected to improve a hospital ward must first obtain a sufficiently faithful representation of what is happening there. A language model can reason over text that already exists; a ward model can reason only over what the ward measured. Sensors—and the structured clinical observations that sometimes substitute for them—are therefore not accessories to clinical AI. They are its sensory layer, and in acute inpatient psychiatry, that layer is thin.

Electronic health records obscure this. A record can be digitally stored, interoperable, and analytically sophisticated while remaining a sparse representation of the physical process of care. It is useful to separate information handling—how observations are stored, connected, and acted upon—from information creation—whether a physical state was measured at all. Digitization and analytics improve the first. They do not cause an unmeasured state to become measured. A nursing note stating that a patient appeared asleep at an observation is machine-accessible information, but it remains an intermittent, observer-dependent sample of a continuous process. This

XSL•FO
RenderX

Viewpoint is about the second problem and about one setting in which it is unusually hard. It is written for the clinicians, nurses, and informaticians who commission, build, or evaluate measurement on acute psychiatric wards, and for engineers designing sensing for them.

The distinction matters because absent measurement is not necessarily neutral missingness. Events that occur when wards are busiest and patients least visible may also be least likely to be documented. A model trained on the resulting record can inherit the ward's blind spots and report them as if they were properties of the patient population. Before asking whether a model is sufficiently capable, clinical AI projects should therefore ask a more primitive question: did the environment generate the signal at all?

In psychiatry, measurement also perturbs what it observes. Monitoring may be experienced as care or surveillance depending on consent, explanation, involuntary status, and the therapeutic relationship. Service users and researchers have raised precisely these concerns about inpatient monitoring technologies [1,2]. A sensory-layer argument that ignores this is an engineering argument in clinical dress.

Inpatient Psychiatry as a Measurement Stress Test

Acute psychiatric wards are not merely under instrumented; they are a measurement stress test because their safety and operational requirements exclude much of the contemporary sensing design space by construction. Clinical states can change across hours while symptom scales, clinical assessment, and routine documentation remain necessarily intermittent. Sleep illustrates the gap. Sleep disturbance is common across psychiatric disorders [3], inpatient sleep is often poor [4,5], and sleep can change repeatedly during an admission. Research actigraphy can provide longitudinal measurement [6-8], but a research device does not thereby become ward infrastructure.

The implementation constraints are structural. Ligature-resistant design is a standard expectation in locked inpatient psychiatric units [9], although the extent to which it is legally required outside hospital settings has been contested [10], making bands, straps, cords, and similar attachments problematic. Rigid or hand-separable components may create self-injury, projectile, or ingestion hazards. Adherence cannot be assumed during disorganization, agitation, refusal, or paranoia; a device that fails when a patient is least well may fail during the period for which it was obtained. In a recent implementation of research actigraphy on a 21-bed adult inpatient psychiatric unit, 68 of 88 patients offered a device accepted it, 1 in 4 refusals cited privacy or surveillance concerns, and wear time and missing data were not reported [11]. Consent frameworks are complicated by involuntary treatment and impaired capacity. Finally, charging, pairing, retrieval, cleaning, contraband control, and troubleshooting consume nursing time. A technically excellent sensor can therefore be operationally unusable.

These constraints expose a general problem for clinical AI. Sensor selection is often treated as a downstream procurement decision made after an analytical use case has been defined. In constrained environments, the measurement architecture is part of the clinical problem itself, and its requirements are of two kinds. Some are necessary conditions: the device must not create a ligature, ingestion, or projectile hazard; it must acquire the signal properties the intended measure depends on; and it must be deployable under the consent conditions of the ward. Others are costs to be minimized once those conditions are met: attachment, power, maintenance, staff time, and data volume. The useful question is therefore not which device measures most accurately but which combination of signal, attachment, power, maintenance, consent, and workflow satisfies the necessary conditions and can remain present long enough to generate a usable trajectory.

The stress test also changes what "failure" means. In ambulatory research, missing data may mainly reduce statistical power; on an acute ward, missingness can coincide with agitation, seclusion, transfer, or other periods of greatest clinical interest. Models can and do learn from patterns of missingness, and informative missingness is itself a signal; but a state that was never measured cannot be recovered from the pattern of its absence, and that is the gap this Viewpoint is about. Reliability therefore includes not only sensor accuracy but also persistence under the conditions that make measurement difficult. A device that is excellent when the patient is cooperative but unavailable during some periods of acute illness can leave clinically important states underrepresented, even when the data that remain are technically clean. Whether that failure is treated as a data quality problem to be repaired afterward, or as a design constraint to be met in advance, is the choice the rest of this Viewpoint is about.

Three Sensing Commitments

There is no single correct sensing architecture for psychiatric care, and the 3 approaches below are not a classification. They are 3 different design commitments—what each asks of the patient, the ward, and the institution—and what each buys in return. A single service may use more than one. None should be treated as the starting point of every measurement problem: where routine care already creates a relevant longitudinal signal, that existing stream should be evaluated before additional sensing is deployed.

The first commits capability to the patient-worn device. Research-grade platforms such as Empatica's EmbracePlus combine accelerometry with multiple physiological channels and a supporting data platform [12]. The commitments are a cooperative wearer, a charging routine, retention of a valuable asset, and band attachment; what they buy is signal richness and longitudinal continuity. This is a strong architecture for research cohorts and ambulatory monitoring where adherence, charging, and device retention are manageable. It can be implemented on an inpatient unit, as a recent report shows [11]; what that report also shows is that acceptance is not universal and persistence is rarely measured.

The second commits to the room rather than the patient. Oxehealth's Oxevision exemplifies fixed ambient sensing in mental health settings, supporting contact-free observations and

XSL•FO
RenderX

sleep- or activity-related measurement and having been evaluated in service improvement and multisite work [13-15]. It removes charging, retrieval, and patient-worn ligature risk, but measurement is bounded by the instrumented room, and deployment becomes a capital and estates decision. Fixed observation in bedrooms also raises distinctive privacy and surveillance concerns. A 2024 systematic review concluded that evidence remains insufficient to establish whether surveillance-based technologies in inpatient and acute mental health settings achieve their intended outcomes, while documenting substantial service-user concerns [1,2].

The third commits to the minimum. Instead of asking how many physiological channels can be added, it asks what minimum measurement is sufficient for a defined clinical or operational purpose and how safely and cheaply it can be acquired across a ward. For sleep, that may mean prioritizing sufficient wrist accelerometry over multimodal fidelity, eliminating charging and retrieval, minimizing rigid and separable components, and treating the device as a consumable. The trade-offs are explicit: less signal richness, no long preadmission personal baseline, continued possibility of refusal or removal, and an environmental burden that generally favors reusable devices on current comparative evidence [16].

The architecture should therefore follow the environment and information requirement. A mature service might use ambient sensing for bedroom observation, focused patient-worn sensing during admission, and richer reusable wearables after discharge; another site may find that routine nursing documentation already performs adequately. The relevant comparison is not sensor versus sensor, but measurement strategy versus the clinical and operational constraints it must survive. This problem-driven approach is consistent with broader recommendations that digital health technology selection begin with the health care problem and implementation context rather than with technology availability alone [17].

This framing also prevents a false contest between “advanced” and “simple” technologies. A multimodal wearable can be the sophisticated answer to a question requiring multiple physiological channels; a room sensor can be the sophisticated answer to an adherence problem; and a deliberately constrained device can be the sophisticated answer to a safety and logistics problem. Sophistication lies in matching architecture to the constraint set, not in maximizing component count. Figure 1 states the rule in 3 steps: define the intended use, establish eligibility against necessary conditions and adequacy, and only then select the least burdensome eligible option, adding nothing the use does not require. Candidates include information streams that routine care already creates. Two deliberately contrasting examples then make the design rule concrete: one asks whether an unusual existing clinical record can make new hardware unnecessary for some purposes; the other asks why existing wrist wearables may still fail as ward infrastructure, and whether a new architecture can be derived from the constraints they do not solve.

XSL•FO
RenderX

Figure 1. Intentional sufficiency as a 3-step design rule. The intended use specifies the derived quantity, acceptable error, and required signal properties. Candidate measurement approaches are then screened for necessary conditions, including safety, consent and acceptability, validity-defining signal properties, and persistence. Among approaches that satisfy these conditions and demonstrate adequate performance, the least burdensome option is preferred. Failed or untested approaches return to step 1 for redesign; untested does not mean failed.

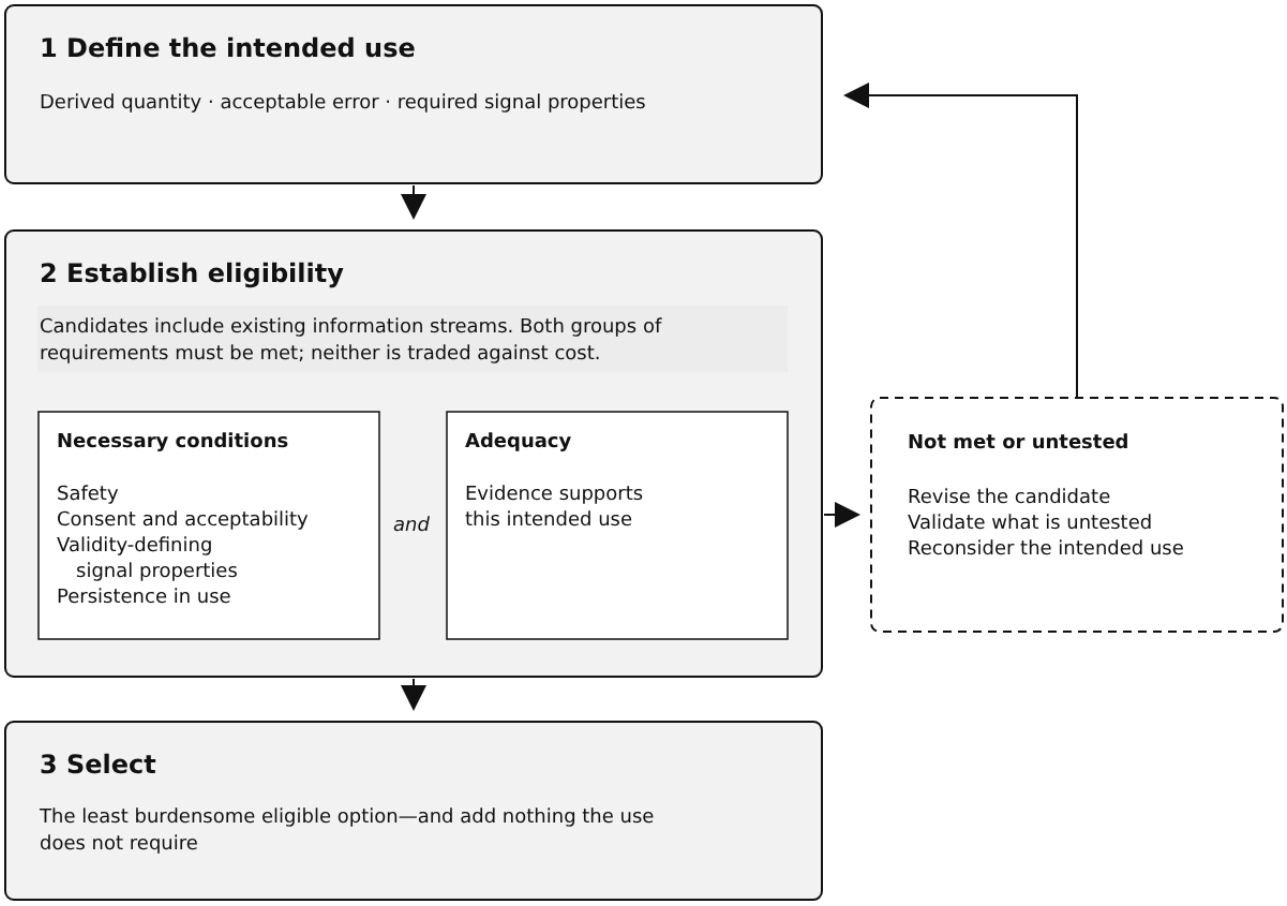


Figure 1 states the rule. Step 1 defines the intended use: the derived quantity the measure will inform, the acceptable error in that quantity, and the signal properties on which the method that produces it depends. Step 2 establishes eligibility. Candidates include information streams that routine care already creates and not only new sensors. Two groups of requirements must both be met, and neither can be traded against cost: necessary conditions (safety, consent and acceptability, preservation of validity-defining signal properties, and persistence in use) and adequacy (evidence that the candidate supports this intended use). A candidate that fails a requirement or whose status is untested leaves the pathway—to be revised, validated, or set aside while the intended use is reconsidered—and reenters only at step 1; untested is not treated as failed. Step 3 selects, among eligible candidates, the least burdensome option and adds nothing the use does not require. The figure states the rule only; it contains no technology, clinical example, or empirical result. How the rule applies to a routine nursing sleep record and to a sensing architecture under development, and the limits of the evidence for each, are set out in Textboxes 1 and 2.

XSL•FO
RenderX

Textbox 1. When information already exists: routine nursing sleep duration records in Taiwan.

Why this example

In the authors' institution, and to our knowledge widely in Taiwan, psychiatric inpatient nursing workflows have long included routine documentation of nightly sleep duration, creating a quantitative longitudinal stream within ordinary care. Nursing observation of sleep is documented in other settings as well, typically as unstandardized narrative notes [18] or within specialist units [19], and both of those reports found imperfect agreement with patient report or actigraphy. What is unusual here is a structured nightly sleep duration figure embedded in routine inpatient nursing workflow and maintained as a system-wide practice for decades; this should not be assumed elsewhere.

What the example tests

If routine care already creates a candidate longitudinal signal, "intentional sufficiency" asks whether it is adequate for the intended use before adding a sensor. In a cross-sectional observational study of 94 acute psychiatric inpatients in Taiwan, nursing estimates of total sleep time were compared with concurrent wrist actigraphy for up to 7 nightly observations early in the admission, not necessarily on consecutive nights. The sources showed moderate association (Pearson r=0.629), good intraclass agreement (intraclass correlation coefficient=0.767), and little mean bias (0.20 hours), but wide limits of agreement (–1.79 to 2.19 hours) and variation by diagnostic group [20]. They should not be treated as interchangeable for individual patients.

What it demonstrates

Existing clinical documentation can be measurement infrastructure rather than being dismissed because it is not sensor-generated. The agreement statistics provide evidence with which a proposed sufficiency boundary can be tested for one use: they indicate how far the nursing figure can be trusted as a proxy for actigraphy-derived total sleep time, at the group level, over a week. They are not a definition of sufficiency, and they do not establish adequacy for every downstream use. Here the 2 design rules diverge. A capability-maximizing rule would add actigraphy to every patient to obtain the richer signal and manage its gaps afterward. Intentional sufficiency asks first whether the existing stream is adequate for the intended use and adds nothing only if that adequacy has been shown for that use. The agreement reported here is evidence for 1 candidate use—group-level description of sleep duration over an admission—and for no other; adequacy for ward-level trajectories or for any individual-level decision has not been established and would have to be tested separately. What makes the existing stream worth testing first is persistence: the nursing record may remain available during refusal, agitation, and compulsory admission, when wearable data may be absent; its completeness and accuracy under those conditions still require evaluation.

What it does not show

The agreement statistics are descriptive and were not designed to answer a hardware specification question. More importantly, the study excluded patients under compulsory admission and enrolled participants only once their condition had improved enough to permit informed consent. The periods in which measurement is most clinically valuable are therefore the periods least represented in the evidence that documentation suffices—an instance of the general problem that missingness on acute wards is not neutral.

Textbox 2. When existing wearables do not fit the ward: a sensing architecture under development.

Why this example

Existing wrist wearables can measure movement well. The problem is that a watch also brings a band, a rigid reusable body, a rechargeable lithium battery, charging, retrieval, cleaning, and an expectation that the patient will keep it on. In an acute psychiatric ward, those features can become safety and workflow liabilities, especially when the patient is most unwell. Our starting question was therefore simple: if we only need longitudinal sleep information, how much of the watch do we actually need?

The window

Three things are easy to conflate here, and the design depends on keeping them apart. The first is sensing capability: how the device samples movement. The second is information extraction: what the device computes from those samples. The third is what the device must store and transmit. Sleep-wake methods commonly reduce movement to summaries over epochs of tens of seconds [6-8]. Established methods may still require sampling at tens of hertz within each epoch, and we do not assume that sampling can be reduced without validation. What can change is the third layer: if one useful summary can leave the device about every 30 seconds, storage and radio use can fall sharply while sensing capability is unchanged. That is the window we are trying to exploit—a reduction in what must be retained and moved, not in what is sensed.

The hardware consequence

Once continuous transmission is no longer assumed, power becomes a different problem. We can ask whether the required monitoring period can be covered without charging, whether a conventional rechargeable lithium battery watch architecture is still necessary, and what that freedom permits in the attachment, enclosure, disposability, and nursing workflow. These choices are coupled: reducing one burden changes what is possible elsewhere in the device.

That is what we are now trying to build. We are testing whether this reduction in data burden is enough to replace the conventional rechargeable watch with a small disposable wrist-worn patch that can remain on the patient for several days without charging, retrieval, or routine staff interaction. We do not yet know whether the trade-offs will work: the device still has to acquire adequate movement data, communicate reliably, remain acceptable and safe on an acute ward, and be inexpensive enough for ward-scale use. The proposed architecture is therefore not evidence that intentional sufficiency works. It defines an experiment that remains to be performed. The concept is patent pending and under development; no claim of accuracy, safety, feasibility, effectiveness, or superiority is made.

XSL•FO
RenderX

Two Demonstrations at Opposite Ends of the Design Space

Intentional sufficiency should do more than redescribe available technologies, so it is worth stating what it decides differently. A capability-maximizing rule selects the source that creates the richest signal and treats loss of data under acute conditions as a quality problem to be repaired afterward—by adherence support, imputation, or exclusion. Such design is not indifferent to feasibility; it treats feasibility as a deployment problem that follows the choice of instrument. Intentional sufficiency reverses the order: it treats persistence under the conditions of the ward as a design constraint that precedes the choice and asks for the least burdensome source that satisfies the intended use. For the same task—a nightly sleep duration trajectory for every patient on an acute ward—the 2 rules can therefore select different sources. The 2 examples below were chosen because they test the rule in opposite directions. The first begins with information that already exists in routine care and asks whether adding a sensor is necessary. The second begins with a measurement gap for which sensors already exist but asks whether their conventional architecture can survive the clinical environment. A useful design rule must be capable of concluding both "do not add hardware" and "hardware must be redesigned," depending on what information exists and what constraints bind.

These are demonstrations of reasoning, not symmetric claims of evidence. The nursing example has empirical paired data and can therefore show what an existing information stream does and does not support. The hardware example is prospective and unfinished: it follows the design logic linearly from a ward problem, through the derived measurement and its epoch structure, to the engineering freedoms that may follow. It does not establish that the resulting device will work. The distinction is deliberate.

Intentional Sufficiency

Reasoning from constraints in this way exposes a design rule, and the rule is separable from any particular device. Intentional sufficiency has 3 parts. First, the intended use frames sufficiency: the investigator specifies the derived quantity, proposes an acceptable error boundary, and identifies the signal properties on which the method depends. These are testable specifications rather than properties supplied by the rule itself. Second, the ward defines feasibility: safety, consent, validity, and workflow are necessary conditions rather than factors to be traded against cost, and together they bound the design space. Third, within that space, the rule selects the least burdensome option that preserves the required properties. The third part is what distinguishes the rule from ordinary task-oriented design: excess capability is not neutral. On an acute ward, it is paid for in attachment, charging, retrieval, and staff time, and ultimately in persistence—the richer device is the one most likely to be absent when the patient is most unwell. Greater sensing capability does not guarantee greater clinically available information when deployment and persistence are constrained.

Agreement with a reference method has a defined place in this structure, and it is a smaller place than our examples might suggest. Where the clinically actionable output is a quantity derived over an epoch—an activity count or sleep-wake classification—sufficiency should be assessed at the level of that derived quantity, and agreement with an already validated method on the same participant and interval is one way of operationalizing the first part of the rule. It is evidence about measurement validity for a stated use. It is not a definition of sufficiency, because the acceptable error depends on the decision the measure will inform, and because a measure can agree well with its reference while being absent for the patients who matter most. This is not permission for arbitrary inaccuracy. It changes where accuracy is optimized: once the derived quantity is adequate for the intended use, additional raw signal fidelity may offer little marginal clinical value while increasing energy demand, cost, rigidity, maintenance, and regulatory burden. In the design exercise of Textbox 2, the question is whether relaxing requirements that the intended derived measure does not need could reduce retention, transmission, or power burdens enough to open otherwise unavailable engineering options. That possibility remains to be demonstrated rather than assumed. How acceptable error is set for a given use is itself an empirical step, not a matter of taste. Two routes are available. Where the downstream method is explicit, its established validation domain defines the properties that must be preserved [21,22]. Where the use is a clinical judgment, acceptable error can be estimated by asking clinicians to make the same decision from alternative subsets of the data, across enough patients and situations to show where decisions begin to change; the number of cases needed is itself uncertain, as Textbox 1 illustrates. Either route makes the threshold a stated, testable quantity rather than a preference.

The engineering consequence can be large even when the clinical quantity is unchanged, as Textbox 2 illustrates. It is the reduction in what must be stored and moved—not in what is sensed—that can change radio duty cycle, mean current demand, and ultimately which power and mechanical architectures are feasible under ward safety constraints.

The principle also has a boundary, and locating that boundary correctly matters more than asserting it. Sufficiency is defined by the downstream analysis rather than by the hardware alone. A signal property may be relaxed only when doing so leaves the intended derived measure within the domain in which its performance has been established. A property that appears unnecessary at the hardware level may still be validity defining for the analytical pipeline that consumes it.

Violating such a validity-defining property can move data outside the domain in which a method was validated without producing an obvious technical failure; the output may remain plausible while becoming invalid. Intentional sufficiency therefore requires explicit identification and empirical testing of which signal properties may be relaxed and which must be preserved. Applied consistently, the principle cuts in both directions: neither fidelity nor constraint should be engineered beyond what the intended measure requires. The target is minimum sufficient fidelity and minimum sufficient constraint, without silent invalidity.

XSL•FO
RenderX

Established fit-for-purpose frameworks ask whether a sensor-based digital clinical measure and its reference method are adequate for a specified context of use [21]. The V3 framework evaluates a biometric monitoring technology through verification, analytical validation, and clinical validation for a defined context of use [22], and V3+ adds usability validation so that the technology also serves its intended users at scale [23]. These frameworks derive requirements from the intended use, as intentional sufficiency does; the difference lies in what is being decided. A validation framework evaluates one candidate at a time and returns a verdict on its adequacy. It does not rank several candidates that all pass, and it does not ask whether the least capable adequate option should be preferred. Intentional sufficiency addresses that remaining decision. Where measurement sources of different capabilities could each be shown to be adequate for the same use, as the routine nursing record and research actigraphy are compared in Textbox 1, validation establishes which are adequate; the third part of the rule then selects among them, retaining additional capability only when its contribution to the intended use justifies the attachment, charging, retrieval, data, and governance burden it adds. The rule therefore adds a comparative design constraint that applies after validation, not an alternative standard of validity: it concerns engineering sufficiency, not validation sufficiency.

Research instrumentation is commonly optimized toward additional channels and higher fidelity. In constrained clinical environments, optimal infrastructure may sometimes move in the opposite direction. That is the central claim of this Viewpoint, and it is narrower than a claim that wearables fail: more sensing capability does not necessarily produce more clinically available information.

Measurement Architecture Includes the Data Path

Neither example ends at the measurement source. A sensor alone does not create an AI-ready ward, and a useful nursing record does not become computational infrastructure merely because it exists. The observation must enter a usable path: acquisition, quality assessment, temporal representation, inference or aggregation, and delivery into an institutional workflow. A proprietary stream requiring manual export simply creates another research silo. The source may be a sensor or a structured clinical observation; in either case, usefulness depends on the full path from capture to interpretation and workflow.

A practical deployment principle is fallback by default. Begin with the lowest-friction source already available at a site, evaluate additional sensing locally in a nonactionable mode, and promote it to actionable use only after adequate local performance is demonstrated. This makes heterogeneity an explicit architectural property and turns the value of added sensing into a testable local question rather than a procurement assumption.

Better measurement also need not first produce bedside predictions. Ward-level measures of changing care load, sleep stability, turnover, or postdischarge burden may be easier to integrate and carry lower costs when wrong than individual treatment alerts. More specific clinical uses can be evaluated separately as evidence accumulates. The purpose of measurement infrastructure is not to fill a database; it is to make previously invisible trajectories available at the level where people and systems can act on them.

This matters for the role of AI. Better measurement does not obligate a system to make increasingly specific predictions about individual patients. The first useful computational layer may simply summarize longitudinal change, identify data quality failure, or expose ward-level trajectories that were previously invisible. Separating measurement from prediction allows the value of the sensory layer to be tested before adding higher-consequence decision support. It also makes failure easier to audit: one can ask separately whether the signal was captured, whether the derived quantity was valid, and whether acting on it improved care.

What Would Have to Be True

The argument fails if more measurement does not improve care. Continuous monitoring elsewhere in medicine shows the risk: reviews have reported that 80% to 99% of monitor alarms may be false or clinically insignificant [24], and one emergency department study found that only 8 of 1049 alarms changed management [25]. The inpatient mental health care evidence is similarly cautionary [1,2]. Agreement with a reference method validates a measurement, not its clinical value. Relevant end points include earlier recognition of deterioration, better-targeted observation, reduced routine intrusion, changes in restrictive practice, workflow, or length of stay. Cheap sensing must also not become a substitute for staffing; its value depends on redirecting human attention rather than simply removing it.

The sufficiency rule also has a methodological limit: the relevant validity domain must be knowable before fidelity is deliberately removed. For analytical pipelines whose signal dependencies are explicit, those dependencies can be specified and tested. For learned models whose dependence on signal properties is opaque or for pipelines expected to change after deployment, the boundary may not be specifiable in advance. In such settings, intentional sufficiency should not be used to justify irreversible loss of signal fidelity; preserving the richer source may be the safer design choice until the downstream dependency is characterized.

Consent must remain substantive under conditions of detention. Any deployment should specify what happens when a detained patient or a patient without capacity declines. A visible, removable device may make refusal more legible than ambient sensing, but that proposition itself requires evaluation with service users. Monitoring technologies cannot earn legitimacy by becoming technically unobtrusive.

Disposability must survive environmental scrutiny. A 2025 systematic review found that 39 of 47 direct comparisons favored reusable devices on carbon footprint, although results varied by device and setting, and methodological heterogeneity precluded quantitative synthesis [16]. The correct comparison is life cycle–based: a consumable must be compared with

XSL•FO
RenderX

reusable alternatives, including decontamination and replacement cycles, and with fixed infrastructure where relevant. This evidence does not currently favor disposability in general; any specific case must be demonstrated rather than assumed.

Finally, the equity case is the least examined and should be tested rather than asserted. Consumable-priced sensing is deployable where capital installation is not, and the World Health Organization Mental Health Atlas 2024 documents severe income-related differences in mental health financing, workforce, and service capacity [26]. If focused disposable sensing has a claim beyond ward economics in well-resourced systems, this is it.

Conclusions: Measure the Ward Before Asking AI to Transform It

AI cannot transform an environment it cannot adequately observe. Acute inpatient psychiatry makes this limitation unusually visible because safety, consent, adherence, and workflow constraints invalidate many assumptions built into conventional sensing. The lesson of this Viewpoint is a design rule for that setting: let the intended use decide what is required, let the ward decide what is feasible, and then add no more sensing than the use requires—because on an acute ward, excess is paid for in persistence. In the Taiwanese example, “adequate” has been examined only at the group level and only in patients who could consent, so it does not yet extend to compulsory admissions or to the periods in which patients were too unwell to enroll; those are precisely the periods a ward-level measure must eventually cover.

Measurement infrastructure should therefore be designed around the clinical environment and the decision-relevant quantity. Rich reusable wearables, ambient sensing, focused consumable sensors, and existing documentation each occupy legitimate regions of the design space. Intentional sufficiency offers one rule for navigating that space: optimize fidelity at the level required by the intended derived quantity, while explicitly protecting the dimensions whose violation creates silent invalidity.

The burden remains empirical. Current evidence is insufficient to assume that monitoring technologies improve outcomes in inpatient mental health care [1,2], and a new sensing architecture earns its place only if it is acceptable to patients and changes what clinicians or health systems can usefully do. Before asking AI to reason about the ward, we first have to give it reliable eyes—and then show that seeing helped.

Acknowledgments

The authors thank the KY Lab at the Institute of Brain Science, National Yang Ming Chiao Tung University, whose hands-on culture of building research tools around the problem at hand helped inspire the design perspective developed in this Viewpoint.

Data Availability

This Viewpoint reports no primary data. No datasets were generated or analyzed. All sources cited are publicly available.

Funding

The authors declared that no financial support was received for this work.

Generative AI Use

Generative AI tools were used during manuscript preparation to assist with literature organization, language editing, and iterative drafting. The authors critically reviewed and verified the manuscript and take full responsibility for its content. The tools used were ChatGPT (GPT-6, OpenAI) and Claude (Claude Fable 5.1, Anthropic). The table-of-contents image was generated with ChatGPT (GPT-6, OpenAI) and edited with Claude (Claude Fable 5.1, Anthropic); it depicts no real patient, ward, or data.

Authors' Contributions

Conceptualization: HMC
Investigation: HMC
Methodology: HMC
Supervision: TBJK
Visualization: HMC
Writing—original draft: HMC
Writing—review and editing: HMC, TBJK

Conflicts of Interest

HMC is named as an inventor on pending patent applications relating to (1) the disposable wearable sensing patch described in Textbox 2 and (2) software architectures for longitudinal clinical decision support, referred to in related work as the Longitudinal Stability Index Platform (LSIP) architecture. No patent has been granted in respect of either application as of the date of submission, and neither is a validated or commercially available product. This Viewpoint discusses named commercial systems (Empatica EmbracePlus; Oxehealth Oxevision) alongside a design approach in which HMC holds an inventorship interest. The authors have no financial or advisory relationship with Empatica or Oxehealth. The design case makes no claim of accuracy, safety, effectiveness,

XSL•FO
RenderX

or superiority. Consistent with Textbox 2, the prototype is presented as an experiment that would test intentional sufficiency, not as evidence that it works. The applications relate to employee inventions made by HMC in the course of his employment at Chang Gung Memorial Hospital. The provisional applications are currently being prosecuted by the inventor, with subsequent ownership and assignment to be handled in accordance with Chang Gung Memorial Hospital's institutional intellectual property arrangements. Neither application is currently licensed or optioned. HMC reports no relevant company equity, officer role, consultancy, or remuneration. TBJK declares no competing interests.

References

1. Griffiths JL, Saunders KR, Foye U, Greenburgh A, Regan C, Cooper RE, et al. The use and impact of surveillance-based technology initiatives in inpatient and acute mental health settings: a systematic review. BMC Med. Nov 29, 2024;22(1):564. [FREE Full text] [doi: 10.1186/s12916-024-03673-9] [Medline: 39614242]
2. Griffiths JL, Saunders KR, Foye U, Greenburgh A, Regan C, Cooper RE, et al. Correction: The use and impact of surveillance-based technology initiatives in inpatient and acute mental health settings: a systematic review. BMC Med. Mar 08, 2025;23(1):141. [FREE Full text] [doi: 10.1186/s12916-025-03979-2] [Medline: 40055676]
3. Freeman D, Sheaves B, Waite F, Harvey AG, Harrison PJ. Sleep disturbance and psychiatric disorders. Lancet Psychiatry. Jul 2020;7(7):628-637. [doi: 10.1016/s2215-0366(20)30136-x]
4. Aboaja AM, Dewa LH, Perry AE, Carey JF, Steele R, Abdelsamie A, et al. Sleep interventions for adults admitted to psychiatric inpatient settings: a systematic scoping review. Sleep Med Rev. Aug 2024;76:101950. [FREE Full text] [doi: 10.1016/j.smrv.2024.101950] [Medline: 38788520]
5. Hughes J, Thomas R, Macklin J, Owen J. Investigating sleep quality on an inpatient psychiatry ward. BJPsych Open. Jun 18, 2021;7(S1):S195. [doi: 10.1192/bjo.2021.525]
6. Plekhanova T, Rowlands AV, Davies MJ, Hall AP, Yates T, Edwardson CL. Validation of an automated sleep detection algorithm using data from multiple accelerometer brands. J Sleep Res. Jun 2023;32(3):e13760. [doi: 10.1111/jsr.13760] [Medline: 36317222]
7. van Hees VT, Sabia S, Jones SE, Wood AR, Anderson KN, Kivimäki M, et al. Estimating sleep parameters using an accelerometer without sleep diary. Sci Rep. Aug 28, 2018;8(1):12975. [FREE Full text] [doi: 10.1038/s41598-018-31266-z] [Medline: 30154500]
8. van Hees VT, Sabia S, Anderson KN, Denton SJ, Oliver J, Catt M, et al. A novel, open access method to assess sleep duration using a wrist-worn accelerometer. PLoS One. Nov 16, 2015;10(11):e0142533. [FREE Full text] [doi: 10.1371/journal.pone.0142533] [Medline: 26569414]
9. Are the recommendations the same for open and/or unlocked psychiatric units? The Joint Commission. URL: https://www.jointcommission.org/en/knowledge-library/support-center/standards-interpretation/standards-faqs/000001231 [accessed 2026-10-14]
10. Finn J, Gooding P, Brophy L, Carlon D, Maylea C. Is anti-ligature an automatic requirement for suicide prevention?: assessing legal obligations in alternative mental health crisis services. Int J Law Psychiatry. 2026;105:102182. [FREE Full text] [doi: 10.1016/j.ijlp.2025.102182] [Medline: 41506150]
11. Culhane BW, Patterson RD, Rahimi-Eichi H, Yip AG, Huesman K, Kostick-Quenet K, et al. Clinical implementation of wearable-derived sleep and activity reporting for inpatient psychiatric monitoring. JMIR Form Res. Sep 21, 2026;10:e88466. [FREE Full text] [doi: 10.2196/88466] [Medline: 42766331]
12. Gerboni G, Comunale G, Chen W, Lever Taylor J, Migliorini M, Picard R, et al. Prospective clinical validation of the Empatica EmbracePlus wristband as a reflective pulse oximeter. Front Digit Health. Dec 4, 2023;5:1258915. [FREE Full text] [doi: 10.3389/fdgth.2023.1258915] [Medline: 38111608]
13. Barrera A, Gee C, Wood A, Gibson O, Bayley D, Geddes J. Introducing artificial intelligence in acute psychiatric inpatient care: qualitative study of its use to conduct nursing observations. Evid Based Ment Health. Feb 2020;23(1):34-38. [FREE Full text] [doi: 10.1136/ebmental-2019-300136] [Medline: 32046991]
14. Ndebele F, Wright K, Gandhi V, Bayley D. Non-contact monitoring to support care in acute inpatient mental health. J Ment Health. Jun 2024;33(3):320-325. [FREE Full text] [doi: 10.1080/09638237.2023.2245882] [Medline: 37589442]
15. Kekic M, Rose A, Bayley D. Enhancing safety in mental health hospitals: a multi-site evaluation of a contactless patient monitoring platform. Ment Health Digit Technol. Jan 22, 2026;3(1):32-45. [doi: 10.1108/MHDT-07-2025-0047]
16. Booth A, Chowaniec M, Goyal S, Faulkner S, Shaw S. The carbon footprints of single-use and reusable medical devices: a systematic review. BMJ Open. Dec 19, 2025;15(12):e108446. [FREE Full text] [doi: 10.1136/bmjopen-2025-108446] [Medline: 41419285]
17. Loo RT, Nasta F, Macchi M, Baudot A, Burstein F, Bove R, et al. Recommendations for successful development and implementation of digital health technology tools. J Med Internet Res. Jun 11, 2025;27:e56747. [FREE Full text] [doi: 10.2196/56747] [Medline: 40499040]
18. Antomás Osés J, Huarte Del Barrio S, Gárriz Murillo C. Nursing registers on sleep and patient perception in a psychiatric unit. A comparative study [Article in Spanish]. An Sist Sanit Navar. Jan 10, 2012;34(3):373-381. [doi: 10.4321/S1137-66272011000300004]

XSL•FO
RenderX

19. Drummond LM, Wulff K, Rani RS, White S, Mbanga-Sibanda J, Ghodse H, et al. How should we measure delayed sleep phase shift in severe, refractory obsessive-compulsive disorder? Int J Psychiatry Clin Pract. Oct 2012;16(4):268-276. [doi: 10.3109/13651501.2012.709866] [Medline: 22809128]
20. Chen HM, Chen YL, Yang CC. Agreement between actigraphy-based sleep measures and routine nursing sleep assessments in acute psychiatric inpatient care. medRxiv. Preprint posted online on September 19, 2026. [doi: 10.64898/2026.09.08.26362317]
21. Bakker JP, McClenahan SJ, Fromy P, Turner S, Peterson BT, Vandendriessche B, et al. A hierarchical framework for selecting reference measures for the analytical validation of sensor-based digital health technologies. J Med Internet Res. Feb 07, 2025;27:e58956. [FREE Full text] [doi: 10.2196/58956] [Medline: 39918870]
22. Goldsack JC, Coravos A, Bakker JP, Bent B, Dowling AV, Fitzer-Attas C, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). NPJ Digit Med. Apr 14, 2020;3:55. [FREE Full text] [doi: 10.1038/s41746-020-0260-4] [Medline: 32337371]
23. Bakker JP, Barge R, Centra J, Cobb B, Cota C, Guo CC, et al. V3+ extends the V3 framework to ensure user-centricity and scalability of sensor-based digital health technologies. NPJ Digit Med. Jan 24, 2025;8(1):51. [FREE Full text] [doi: 10.1038/s41746-024-01322-2] [Medline: 39856145]
24. Cvach M. Monitor alarm fatigue: an integrative review. Biomed Instrum Technol. 2012;46(4):268-277. [doi: 10.2345/0899-8205-46.4.268] [Medline: 22839984]
25. Fleischman W, Ciliberto B, Rozanski N, Parwani V, Bernstein SL. Emergency department monitor alarms rarely change clinical management: an observational study. Am J Emerg Med. Jun 2020;38(6):1072-1076. [doi: 10.1016/j.ajem.2019.158370] [Medline: 31387812]
26. Mental Health Atlas 2024. World Health Organization. Sep 2, 2025. URL: https://www.who.int/publications/i/item/9789240114487 [accessed 2026-09-22]

Edited by T de Azevedo Cardoso; submitted 22.Sep.2026; peer-reviewed by Y Wang, S Gupta, D Cook; comments to author 13.Oct.2026; revised version received 31.Oct.2026; accepted 04.Nov.2026; published 19.Nov.2026

Please cite as:
Chen H-M, Kuo TB-J
Measure the Ward First: Intentional Sufficiency as a Design Rule for Sleep Sensing in Acute Inpatient Psychiatry
J Med Internet Res 2026;28:e112674
URL: https://www.jmir.org/2026/1/e112674
doi: 10.2196/112674
PMID:

©Hong-Ming Chen, Terry B-J Kuo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX