The ILHBN: challenges, opportunities, and solutions from harmonizing data under heterogeneous study designs, target populations, and measurement protocols

Sy-Miin Chow,[1,] Inbal Nahum-Shani,[2] Justin T. Baker,[3,4] Donna Spruijt-Metz,[5,6]
Nicholas B. Allen,[7] Randy P. Auerbach,[8] Genevieve F. Dunton,[5,9] Naomi P. Friedman,[10,11]
Stephen S. Intille,[12,13] Predrag Klasnja,[14,] Benjamin Marlin,[15] Matthew K. Nock,[16,17,18,19]
Scott L. Rauch,[3,4] Misha Pavel,[12,13] Scott Vrieze,[20] David W. Wetter,[21,22] Evan M. Kleiman,[23]
Timothy R. Brick,[1,24] Heather Perry,[1] Dana L. Wolff-Hughes,[25] Intensive Longitudinal Health
Behavior Network (ILHBN)

[1]Department of Human Development and Family Studies, The Pennsylvania State University, University Park, PA 16802, USA
[2]Institute for Social Research, University of Michigan, Ann Arbor, MI, USA
[3]Department of Psychiatry, McLean Hospital, Boson, MA, USA
[4]Department of Psychiatry, Harvard Medical School, Boson, MA, USA
[5]Department of Psychology, University of Southern California, Los Angeles, CA, USA
[6]Center for Economic and Social Research, University of Southern California, Los Angeles, CA, USA
[7]Department of Psychology, University of Oregon, Eugene, OR, USA
[8]Department of Psychiatry, Columbia University, New York, NY, USA
[9]Department of Population and Public Health Sciences, University of Southern California, Los Angeles, CA, USA
[10]Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, CO, USA
[11]Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, CO, USA
[12]Khoury College of Computer Sciences, Northeastern University, Boston, MA, USA
[13]Bouvé College of Health Sciences, Northeastern University, Boston, MA, USA
[14]School of Information, University of Michigan, Ann Arbor, MI, USA
[15]College of Information and Computer Sciences, University of Massachusetts Amherst, Amherst, MA, USA
[16]Department of Psychology, Harvard University, Cambridge, MA, USA
[17]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA
[18]Franciscan Children's, Boston, MA, USA
[19]Children's Hospital, Boston, MA, USA
[20]Department of Psychology, University of Minnesota, Minneapolis, MN, USA
[21]Department of Population Health Sciences, University of Utah, Salt Lake City, UT, USA
[22]Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA
[23]Department of Psychology, Rutgers, The State University of New Jersey, Piscataway, NJ, USA
[24]Institute for Computational and Data Sciences, The Pennsylvania State University, University Park, PA, USA
[25]National Cancer Institute, National Institutes of Health, Bethesda, MD, USA
Correspondence to: S-M Chow, quc16@psu.edu

ABSTRACT

The ILHBN is funded by the National Institutes of Health to collaboratively study the interactive dynamics of behavior, health, and the environment using Intensive Longitudinal Data (ILD) to (a) understand and intervene on behavior and health and (b) develop new analytic methods to innovate behavioral theories and interventions. The heterogenous study designs, populations, and measurement protocols adopted by the seven studies within the ILHBN created practical challenges, but also unprecedented opportunities to capitalize on data harmonization to provide comparable views of data from different studies, enhance the quality and utility of expensive and hard-won ILD, and amplify scientific yield. The purpose of this article is to provide a brief report of the challenges, opportunities, and solutions from some of the ILHBN's cross-study data harmonization efforts. We review the process through which harmonization challenges and opportunities motivated the development of tools and collection of metadata within the ILHBN. A variety of strategies have been adopted within the ILHBN to facilitate harmonization of ecological momentary assessment, location, accelerometer, and participant engagement data while preserving theory-driven heterogeneity and data privacy considerations. Several tools have been developed by the ILHBN to resolve challenges in integrating ILD across multiple data streams and time scales both within and across studies. Harmonization of distinct longitudinal measures, measurement tools, and sampling rates across studies is challenging, but also opens up new opportunities to address cross-cutting scientific themes of interest.

Lay Summary

Health behavior changes, such as prevention of suicidal thoughts and behaviors, smoking, drug use, and alcohol use; and the promotion of mental health, sleep, and physical activities, and decreases in sedentary behavior, are difficult to sustain. The ILHBN is a cooperative agreement network funded jointly by seven participating units within the National Institutes of Health to collaboratively study how factors that occur in individuals' everyday life and in their natural environment influence the success of positive health behavior changes. This article discusses how information collected using smartphones, wearables, and other devices can provide helpful active and passive reflections of the participants' extent of risk and resources at the moment for an extended period of time. However, successful engagement and retention of participants also require tailored adaptations of study designs, measurement tools, measurement intervals, study span, and device choices that create hurdles in integrating (harmonizing) data from multiple studies. We describe some of the challenges, opportunities, and solutions that emerged from harmonizing intensive longitudinal data under heterogeneous study and participant characteristics within the ILHBN, and share some tools and recommendations to facilitate future data harmonization efforts.

Implications

Practice: Understanding and maintenance of healthy lifestyles can benefit from harmonizing intensive longitudinal data (ILD) across studies that may show heterogeneity in study design, device choices, and data types.

Policy: Policymakers who want to increase and maintain positive health behavior changes should support ILD harmonization initiatives under heterogeneous designs to enhance the quality and utility of expensive and hard-won ILD.

Research: Future research should explore ways to integrate and harmonize ILD across studies under heterogeneous designs to provide comparable views of data from different studies and amplify scientific yield.

Data harmonization provides tremendous opportunity for advancing research. Across disciplines, several harmonization efforts are underway to integrate data across scientific contexts and the lifespan [1, 2]. Recent initiatives [3] have also underscored the importance of using big data, especially intensive longitudinal data (ILD), with innovative computational and modeling approaches to develop dynamic health behavior theories. However, harmonizing ILD poses unique challenges. In this article, we discuss some of the data harmonization opportunities, challenges, and solutions from the ILHBN.

Established through support of the National Institutes of Health, the ILHBN sought to use ILD to collectively: (a) understand and intervene on behavior and health and (b) develop new analytic methods to innovate behavioral theories and interventions. The ILHBN's seven study projects were served by a Research Coordinating Center (RCC) and differed in their target health behaviors, study designs, populations, and participant inclusion criteria (see Table 1)

At its inception, the ILHBN engaged in discussions to find opportunities and consensus for data harmonization. Interest and success have been fueled by two key motivators: unique problems posed by use of distinct longitudinal measures, measurement tools, and sampling rates by different studies; and shared scientific interests in cross-cutting themes of interest to multiple projects. Next, we describe some of the harmonization opportunities and solutions that emerged from the network's collaborative efforts.

INCLUSION OF NEW ITEMS TO HARMONIZE ECOLOGICAL MOMENTARY ASSESSMENT DATA

The ILHBN studies targeted processes that unfolded across a multitude of time scales (Fig. 1), with data collection spanning 3 months to 4 years. Ecological momentary assessment (EMA) was a key component of all seven studies, with five studies (Temporal Influences of Movement and Exercise [TIME], Dynamic Models of Behavior [DMB], Mobile Assistance for Regulating Smoking [MARS], Sensing and Mobile Assessment

in Real Time Study (SMART), and Mobile Assessment for the Prediction of Suicide [MAPS]) using variations of the multiple burst design [4]. Each burst period lasted between 4 to 10 days, with 5 to 6 measurements per day. Most ILHBN studies also used "diary" designs, collecting daily or biweekly self-reports over one or more years.

Adopting similar measures at the inception of a new network was challenging, but offered a more direct and efficient method for data harmonization than other post-hoc alternatives [5, 6]. However, full standardization of all study-related measures was not always feasible. An alternative approach the ILHBN adopted was to standardize when sites have overlap in planned measures. For example, affect was a common construct across multiple studies. Participating Principal Investigators (PIs) agreed to include seven common EMA affect items (i.e., relaxed, tense, energetic, fatigue, happy, sad, and stressed). Doing so allowed researchers to identify the scales of common unobserved (latent) constructs (e.g., positive affect) and study-specific items relative to those of the common items [2].

Despite the use of common EMA items, the instructions, wording, and measurement scales were not standardized. For instance, studies with multiple measurements within a burst design asked participants, "*Right now*, how stressed do you feel?" while studies adopting a daily or weekly design, asked participants, "*Over the past day (week)*, how stressed did you feel?" In addition, whereas most of the studies adopted a 5-point ordinal scale (from 1 = not at all to 5 = extremely) for EMA affect items, MAPS used a continuous sliding scale to increase the variability of participants' responses. Still, the availability of common or parallel measures created opportunities to test the constraints of data harmonization within the realm of ILD.

TOOLS TO FACILITATE WITHIN- AND CROSS-STUDY INTEGRATION OF TIME SCALES

Effective observational and interventional research requires understanding and consideration of the dynamic variations of health behaviors [4]. Across studies, the differences in

Table 1 | Summary of the seven U01 studies within the ILHBN

Study names	MAPS[a]	SMART[b]	BLS[c]	TIME[d]	DMB[e]	MARS[f]	CoTwins[g]
Sponsoring NIH institutes	NIMH	NIMH	NIMH	NHLBI/OBSSR	NCI/OBSSR	NCI	NIDA
Projected sample size (N)	200	600	100	246	120	112	1,558[h]
Target age range of participants	13–18 years old	18 + (adults); 12–17 (adolescents)	18 years and older	18–29 years old	18 years and older	18 or older	14–24
Geographical coverage	New York, Pittsburgh	Massachusetts	Massachusetts	U.S. National	California, Washington	Utah	Mostly in Colorado
Key health behaviors	Suicidal thoughts and behaviors (STBs)	Suicidal thoughts and behaviors (STBs); sleep	Sleep; episodes of depression, mania, or psychosis	Physical activity, sedentary behavior; sleep	Physical activity; sedentary behavior	Cigarette use	Substance use
Participant inclusion criteria[i]	Three groups, all with primary mood, anxiety, and/or substance use disorder, but categorized as: (1) *Suicide Ideators*: (score ≥ 4 Beck's Scale for Suicide Ideation [24, 25]; (2) *Suicide Attempters* with current suicidal ideation and suicide attempt in the past 12 months; (3) *Psychiatric Controls* with no lifetime history of suicidal thoughts and behaviors	Adults and adolescents with presentation of a problem that included suicidal behaviors or severe suicidal ideation, recruited from the Acute Psychiatry Service and children's Inpatient Psychiatry Service of affiliated hospitals	Two groups with a psychotic illness and symptoms such as hallucination, delusions, conceptual disorganization, or negative symptoms, categorized based on the Structured Clinical Interview for DSM-V [26] into a: (1) *Non-affective psychosis group* consisting of individuals with schizophrenia, psychosis, and delusional disorders; (2) *Affective psychosis group* consisting of individuals with major or bipolar depression and psychosis	Participants with intention to engage in recommended levels of moderate-to-vigorous physical activity (≥150 min/week moderate or ≥75 min/week vigorous intensity) within the next 12 months	Overweight, (body mass index between 25–45), sedentary (as guided by International Physical Activity Questionnaire [27] healthy or Type-2 diabetic adults (under medications to control glucose with approval of a healthcare provider) able to participate in and track physical activity	Current smoker with an average of at least 3 cigarettes per day motivated to quit within the next 30 days; valid home address; at least marginal health literacy (>45 on the Rapid Estimate of Adult Literacy in Medicine [28]	Twin pairs born in Colorado from 1998 to 2002, and new enrollment of additional twins born in Colorado currently attending high school.

Table 1 | Continued

Study names	MAPS[a]	SMART[b]	BLS[c]	TIME[d]	DMB[e]	MARS[f]	CoTwins[g]
Selected health behavior predictors	Emotional distress; social dysfunction; sleep disturbance	Data-driven search	Processing speed; cognitive control	Affect; productivity, demands; stress, attention; self-control; coping; emotional regulation; habit; intentions; hedonic motivation; deliberation	Commitment; goal; self-efficacy; intrinsic and extrinsic motivation; social support; personality traits	Affect; self-regulation; context; smoking outcome expectancies; urge, self-efficacy; and engagement	Affect; behavioral disinhibition; personality; context and environment; cognitive function including executive functioning

[a] The Mobile Assessment for the Prediction of Suicide (MAPS; PIs: Auerbach & Allen) leveraged data from adolescents' naturalistic use of smartphone technology over 6 months to improve the short-term prediction of suicidal behaviors among high-risk adolescents, 13–18 years old.

[b] The Sensing and Mobile Assessment in Real Time study (SMART; PI: Nock), monitored adolescents (12–17 years) and adults (18+ years) to understand and improve suicide prediction during the highest risk period—the 6 months after a hospital visit for the treatment of suicidal thoughts.

[c] The Bipolar Longitudinal Study (BLS; PIs: Baker & Rauch) examined early biological, environmental, and social factors that triggered mania and psychosis in at-risk individuals (≥18 years) over 48 months to identify who and why individuals were at elevated mental health risk.

[d] The Temporal Influences of Movement and Exercise (TIME; PIs: Dunton & Intille) Study utilized mobile ILD from young adults (18–29 years) across 12 months to elucidate the mechanisms underlying physical activity adoption versus maintenance, sedentary time, and sleep.

[e] The Dynamic Models of Behavior (DMB) Study (PIs: Spruijt-Metz, Klasnja, & Marlin) was a micro-randomized trial (MRT) to devise effective and scalable solutions for helping overweight but otherwise healthy adults (≥18 years) to sustain physical activity and limit sedentary time in ways that adapted to their ongoing dynamic changes.

[f] The Mobile Assistance for Regulating Smoking (MARS; PIs: Nahum-Shani, & Wetter) Study employs a MRT over 10 days on 112 smokers (≥18 years) attempting to quit to inform real-time, real-world self-regulatory strategies.

[g] The Colorado Online Twins Study (CoTwins; PIs: Vrieze, & Friedman) assessed, over 36 months, how individual attributes interacted with environmental and social context to affect normative developmental trends, substance use, and addiction in twins between 14 and 24 years old.

[h] This is the actual N, where 898 are twins and the remainder are the twins' parents.

[i] Other inclusion criteria not spelled out include language fluency requirements, smartphone ownership, or sufficient access to technology (e.g., computer, tablet) to enable provision of the ILD targeted by the study.

Fig 1 | An overview of the designs of the seven network studies. Line or point density within the block arrow indicates response frequency. EMA = Ecological momentary assessment; Exec func = Executive functioning tasks; MRT = Micro-randomized trial.

measurement intervals and total time lengths reflected the studies' distinct priorities in capturing the dynamics of relevant phenomena while maintaining participant engagement over the entire study span. For instance, in MARS, the micro-randomized trial was employed 3 days prior to the participants' quit date through 7 days post-quit to capture the period of high risk for relapse [7]. In contrast, the TIME study interspersed 24 bursts throughout a year (i.e., approximately twice per month) to obtain snapshots of momentary states for predicting participants' physical activity and sedentary behavior [8].

To resolve between-study differences in measurement time scales, we recommend recording precise timestamps and anchoring the data to a common time zone to allow a standardized way of computing the time elapsed between successive occasions—a critical feature in most ILD. For instance, individuals' time scales of recovery from extreme affective states can help distinguish normative from problematic health outcomes [9]. To this end, continuous-time modeling tools, such as software packages developed by members of the RCC [10, 11], provide options to accommodate diversely and irregularly spaced data (see Table 2).

The network has also developed tools that help visualize and explore multiple streams of ILD (and hence, time scales). The Deep/Digital Phenotyping Dashboard (DPdash; [12]) is a web-viewable, customizable dashboard to facilitate visualization of continuously acquired data from multiple data streams (e.g., actigraphy, magnetic resonance imaging data, missing data patterns, and other data collection issues).

Members of the TIME study [13, 14] have developed a customized tool, Signaligner Pro, that can provide simultaneous visual display and algorithms that enable classification of different types of physical activities, sleep/wear/non-wear

statuses [15], and segments of problematic data that warrant closer investigation or possible data removal. Tools such as DPdash and Signaligner Pro (see Table 2 for a full list) provide helpful exploratory information to identify appropriate time scales for within- (and cross-) study harmonization of time scales across data streams (e.g., EMA and passive sensor data).

DATA PRIVACY, SPATIAL, AND TIME HETEROGENEITY IN HARMONIZING LOCATION DATA

Location data collected by phones and wearable devices using Global Positioning System (GPS) and other technologies used with mapping databases can provide rich contextual information about neighborhood characteristics, as well as where, when, and with whom individuals experience conditions that may impact health-related outcomes. However, intensive, timestamped geolocation data, even with other identifiers removed, may still reveal potentially sensitive and identifying information if the data were shared with other entities. For instance, prolonged stays at specific locations at nighttime could expose individuals' residential addresses; repeated routes could allow reconstruction of the participants' regular travel trajectories and increase the participants' vulnerability for identity exposure. To avoid data privacy issues from sharing raw geolocation data, the network opted to harmonize and share derived measures such as distances and time spent at categories of landmarks (e.g., recreation parks, tobacco retail outlets); neighborhood characteristics such as crime rate; and markers of mobility patterns, such as activity space, namely, the typical mobility area of an individual over some specified time frame [16].

Table 2 | Challenges, solutions, and tools developed

Challenges	Solutions/Tools for Resolving Challenges	Where to Access Tools
Challenges in collecting ILD and/or integration with data inspection and analysis	The Beiwe digital phenotyping pipeline [29] includes: Beiwe platform for high-throughput data collection, Forest library for data analysis, and Tableau for interactive data visualization.	www.beiwe.org
	DPdash [12] is a web-viewable dashboard that allows users to visualize continuously acquired dynamic data coming in from multiple data streams that can include smartphone-based passive and active data, surveys, actigraphy, Magnetic Resonance Imaging, and derived variables. DPdash allows users to customize what data to display, how they are displayed, and ways to flag missing data patterns and other data collection and compliance issues.	https://github.com/AMP-SCZ/dpdash
	Study Pay is an app that helps monitor and manage participant compensation and compliance. A customized dashboard (called the "Hub") is available for researchers to visualize and obtain other summary statistics on participant data, the form of compensations, as well as utility functions to automate delivery of participant compensation. Study Pay can be downloaded by the participants as a mobile app to their phones and used to track the compensations received to date.	https://play.google.com/store/apps/details?id=edu.harvard.fas.nocklab.studypaymobile&hl=en_US&gl=US
	Grafana is a tool that allows us to monitor the number of EMAs delivered, compliance, sensor battery life and connectivity, and cell phone battery life. The tool can be used to monitor EMA compliance and helps staff and participants troubleshoot equipment issues that may be occurring.	https://grafana.com/get/?pg=hp&plcmt=what-is-graf-btn-2&cta=download
	The Signaligner Pro [13] provides simultaneous visual display as well as annotations of possible physical activity statuses at the individual level (e.g., wake, sleep, walking, running, biking). Signaligner Pro is pre-packaged with three algorithms that enable classification of different types of physical activities, sleep/wear/non-wear statuses, as well as segments of problematic data that warrant closer investigation or possible data removal.	https://crowdgames.github.io/signaligner/
	The Wear-IT framework [30] is a framework for adaptive real-time scientific data collection and intervention using commodity smartphones and wearable devices	https://sites.psu.edu/realtimescience/projects/wear-it/
	Lochness is a data management tool designed to periodically poll and download data from various data archives into a local directory, often referred to as a "data lake." Lochness natively supports pulling data from Beiwe, MindLAMP, XNAT, REDCap, Dropbox, Box, external hard drives, and more.	https://github.com/AMP-SCZ/lochness
Lack of tools for collecting and integrating study-specific digital health behavior measures and passive data from smartphones	Effortless Assessment Research System (EARS) tool is a suite of programs that collect multiple indices of a person's social and affective behavior passively, via their naturalistic use of a smartphone [31]. EARS places a unique emphasis on capturing and deciphering affective, social, and contextual cues from multiple forms of social communication on the phone. Signals collected include facial expressions, acoustic vocal quality, natural language use, physical activity, music choice, and geographical location	https://www.ksanahealth.com/ears-research-platform
Lack of tools for automated processing of high volumes of intensively measured location data	GPS2space [17], is a free and open-source Python library to facilitate the processing of GPS data, integration with GIS to derive distances from landmarks of interest, as well as extraction of two spatial features: activity space of individuals and shared space between dyadic members.	https://gps2space.readthedocs.io/en/latest/

Table 2. Continued

Challenges	Solutions/Tools for Resolving Challenges	Where to Access Tools
Challenges in extracting behavioral features/patterns from GPS data without releasing/relying on identifiable information	The Deep Phenotyping of Location (DPLocate) is an open-source GPS processing pipeline that securely reads encrypted GPS coordinates, uses temporal filtering to detect temporal epochs with statistically acceptable data points and then uses density-based spatial clustering algorithm to identify the most visited places by the individual during the longitudinal study as points of interest (POI). Times of the day are allocated to the POIs with emphasis on behaviorally meaningful Time-Bands for sleep, socializing, work/school and nightlife. The repeating patterns of behavior help researchers identify phenotypes without getting direct access to the identifiable data.	https://github.com/dptools/dplocate
Difficulties in implementing sustainable JIT interventions in everyday life	The HeartSteps mobile app/dashboard [32] is a platform utilized by the DMB study to implement JITAIs that encourage walking and breaks in sedentary activity, as well as monitoring adherence to the intervention designs. Currently paired with the Fitbit, the mobile app helps individuals set activity goals, and plan how they maintain the motivation to remain physically active in their daily lives.	https://heartsteps.net
Lack of tools for analysis of ILD	Dynamic Modeling in R (dynr) is an R package for modeling linear and nonlinear dynamics with possible sudden (regime) shifts	https://dynr.github.io/
Challenges in harmonizing accelerometer data collected with different devices	The MIMS-unit algorithm is developed to compute Monitor Independent Movement Summary Units, a measurement to summarize raw accelerometer data while ensuring harmonized results across different devices.	https://rdrr.io/cran/MIMSunit/
Challenges in extracting sleep parameters from raw longitudinal accelerometer data	The Deep Phenotyping of Sleep (DPSleep) is an open-source pipeline to translate raw accelerometer data to minute-based activity scores, extract sleep epoch and sleep quality parameters, perform data visualizations to facilitate quality control and detection of idiosyncratic behaviors such as no sleep, time zone shift and disrupted sleep.	https://github.com/dptools/dpsleep

GPS2space (1) [17] is a free and open-source Python library developed by the RCC to help facilitate the processing of GPS data, integration with Geographical Information System data to derive distances from landmarks of interest, and extraction of individuals' activity space and shared space—the proportion of overlap in activity spaces between any two individuals. The team used GPS2space with data from the CoTwins study to explore seasonal changes in individuals' activity space and twin siblings' shared space, and corresponding gender and zygosity differences [18]. Thus, through use of shared tools such as GPS2space, the network was able to harmonize a list of derived spatial features.

In selecting an aggregation time scale to derive spatial summary measures, we recommend extracting spatial features at the finest possible time intervals that can be afforded by the study's computational power and resources to preserve nuanced changes. Aggregation to coarser time scales (e.g., from hourly to daily) can be pursued later as needed. We also recommend that data processing procedures for preparing derived measures be standardized whenever feasible, and clearly documented to facilitate replicability of modeling results. Examples of standardization considerations include procedures for outlier identification and treatment, choice of spatial reference system, procedures and tuning parameters used to derive spatial features, and the time granularity of the shared data.

HARMONIZING ACCELEROMETER DATA ACCOUNTING FOR DEVICE-RELATED HETEROGENEITY

Advances in wearable sensors have enabled lower costs and greater use of accelerometer data in health behavior research. However, raw acceleration data are only available in some devices with most consumer devices providing metrics such as steps inferred from accelerometer data using proprietary algorithms. These devices rarely provide accurate estimates of energy expenditure in free-living environments, showing greatly reduced sensitivity (power) at detecting and differentiating low-intensity daily activities [19].

Within the ILHBN, there was considerable variability in the devices used (see Table 3). This variability presented challenges for direct harmonization and underscored the need to explore device-independent summary statistics with good sensitivity and specificity in physical activity. The Monitor-Independent Movement Summary (MIMS) units is a measure found to yield less inter-device variability in controlled laboratory settings [20]. Computation of MIMS begins with interpolation and extrapolation of raw acceleration data to circumvent device differences in dynamic g-range (the greatest amount of acceleration that can be measured accurately by a device, typically in ±g) and sampling rate, followed by bandpass filtering to remove data outside of the range of meaningful human movement. Data aggregation is then performed across a user's specified time scale to produce a measure of activity intensity. MIMS-units have been used in large-scale studies to summarize accelerometer data and derive population-based comparison metrics for overall activity levels, but properties of MIMS in free-living environments are still an ongoing area of research [21]. The massive volumes of accelerometer data collected in diverse and free-living environments in the ILHBN gave rise to new harmonization

Table 3 | Selected sensor metadata from the U01 studies

Device	Feature	Project						
		MAPS	SMART	BLS	TIME	DMB	MARS	CoTwins
Phone	Platform	EARS[a]	Beiwe[b], LifeData[c]	Beiwe	Custom software	HeartSteps[d]	mCerebrum[e]	Custom software
	OS/Device	iOS + Android	iOS + Android	iOS + Android	Android	iOS + Android	Android	iOS + Android
	Device owner	Participant	Participant	Participant	Participant	Participant	Study	Participant
	Accelerometer sampling rate	50 Hz	10 Hz	10 Hz	~50 Hz	–	250Hz	–
	Location sampling rate	~Every 15 min[f]	~Every 10 min[f]	~Every 2–10 min[f]	~Every minute[f]	~Every 5 min[f]	Every 30 s	Every 5 min on Android; iOS[f]
Wearable	Device	–	Empatica Embrace	GENEactiv Original	Fossil Sport Gen 4/5[f]	Fitbit Versa	MotionSense	–
	Charging	–	Once per day	Monthly	Once per day	Once per day	Once per day	–
	Location (wrist)	–	Dominant	Nondominant	Wrist	Nondominant	Both	–
	Accelerometer sampling rate	–	32 Hz	20 Hz	~50 Hz	–	250 Hz	–
	Gyroscope sampling rate	–	–	–	–	–	250 Hz	–
	Physio sampling rate	–	1 Hz (temperature); 4 Hz (EDA)	20 Hz	–	0.2 Hz	200 Hz	–
	Accelerometer g range	–	±16 g	±9 g	±8 g	–	±2 g	–

Physio = Physiological data; iOS=iPhone operating system; EDA = Electrodermal activity;

[a] See Ref. [31];
[b] See https://www.hsph.harvard.edu/onnela-lab/beiwe-research-platform/; Ref. [29].
[c] See https://www.lifedatacorp.com/.
[d] See Ref. [32];
[e] See Ref. [33].
[f] Varies by device or across participants.

opportunities and underscored the need for new methodological innovations.

UNDERSTANDING AND SUSTAINING PARTICIPANTS' ENGAGEMENT IN ILD STUDIES

Participant retention and instances of complete data almost always decline over time in EMA studies, even when monetary incentives are offered. To date, limited attention has been given to understanding the concept and dynamics of engagement in digital data collection [7, 22]. One cross-cutting theme that has fueled harmonization efforts within ILHBN resided in the studies' shared interests in understanding characteristics that shape engagement in EMAs. To address these questions, the engagement subgroup, led by [MASKED], initiated a project to evaluate the dynamics of self-report completion in the network studies. Participating sites worked collaboratively to select key constructs to clarify *how past EMA history predicts future EMA completion*, as modulated by factors such as time of day, day of week, in-study time length, observational versus intervention study, and study population. Harmonization of paradata (data about when and how participants complete self-report questions) is underway to delineate participants' EMA completion patterns over time, and risk for non-ignorable missingness [23].

Other helpful solutions included Study Pay, an app developed by the SMART Study for monitoring and managing participant compensation and engagement, with a customized dashboard for visualizing and monitoring participants' summary statistics (e.g., number of completed surveys), compensation history, modality, and delivery (see Table 2).

CONCLUSION

There has been clear consensus within the ILHBN that caution should be exercised when deciding whether and how to homogenize data across studies and that harmonization efforts should be guided by clear scientific questions. This vision facilitated the design and sharing of tools to enhance the network's ILD harmonization efforts. The ILHBN has been sharing computational and modeling resources, code, and tutorials at the network website at [WEBSITE MASKED]. Preparations of machine-readable metadata, protocol descriptions, data processing steps, and well-annotated scripts are imperative to ensure successful sharing and interpretations of results using the harmonized data.

Acknowledgments

Compliance with Ethical Standards

Conflict of Interest: Research reported in this publication was supported by the NIH under the ILHBN Cooperative Agreement Program. Travel support from the grants awarded to the seven U01 studies and one U24 research coordinating center has made possible some of the collaborative harmonization efforts discussed in the manuscript. Dr. Allen is employed by and has an equity interest in Ksana Health Inc., and has been a paid advisor to Google Health and Snap Inc. In addition, Dr. Nock has been a paid consultant in the past year for Microsoft Corporation, the Veterans Health Administration, and Cerebral; an unpaid scientific advisor for Empatica, Koko, and TalkLife; and receives publication royalties from Macmillan, Pearson, and UpToDate. Dr. Pavel has been a paid consultant for Honeywell Research Laboratory, Brno, CR, and is a member of Data Monitoring Committee of the "Understanding America Study." Dr. Auerbach is an unpaid scientific advisor for Ksana Health, Inc. Dr. Rauch has been employed by Mass General Brigham/McLean Hospital; paid as secretary of Society of Biological Psychiatry, and for Board service to Mindpath Health/Community Psychiatry and National Association of Behavioral Healthcare; served as volunteer member of the Board for Anxiety & Depression Association of America, and The National Network of Depression Centers; received royalties from Oxford University Press, American Psychiatric Publishing Inc, and Springer Publishing; received research funding from NIMH.

Human Rights: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee of each study and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The MARS Study's IRB approval has been obtained from the University of Utah under IRB approval number IRB_00112287; The BLS IRB approval has been obtained from the McLean Hospital and Harvard Medical School under IRB approval number 2015P002189; The DMB Study's IRB approval has been obtained from the University of Southern California under IRB approval number UP-18-00791 and from the University of Michigan under IRB approval number HUM00196881; The MAPS Study IRB approval has been obtained from Columbia University/New York State Psychiatric Institute under IRB#7630; The TIME Study IRB approval has been obtained from the University of Southern California under IRB approval number HS-18-00605; The CoTwins Study IRB approval has been obtained from the University of Colorado Boulder under IRB approval number IRB14-0433; The SMART Study IRB approval has been obtained from Harvard University under IRB approval number IRB18-1749;

Informed Consent: Informed consent was obtained from all individual participants included in the study.

Welfare of Animals: This article does not contain any studies with animals performed by any of the authors.

Transparency statements:

1.**Study registration:** Although this study was not formally registered, Projects MARS and DMB, which were described in the article, are registered at Clinicaltrials.gov under the titles

Project MARS Mobile-Assistance for Regulating Smoking and HeartSteps, respectively.

2. **Analytic plan pre-registration**: Analysis and summaries described in the article were not formally registered, however some of the studies in the Network have been pre-registered including Project MARS and Project DMB.

3. **Data availability**: All metadata for the seven studies described in the manuscript have already been summarized in the manuscript as Tables and Figures. Further details can be obtained by emailing the corresponding author.

4. **Analytic code availability**: There is no analytic code associated with this study.

5. **Materials availability**: Materials used to conduct the study are not publicly available.

REFERENCES

1. Ruggles S. The Minnesota population center data integration projects: Challenges of harmonizing census microdata across time and place. Paper presented at: *Proceedings of the American Statistical Association, Government Statistics Section*. Alexandria, VA: American Statistical Association; 2014:1405–1415.

2. Curran PJ, Hussong AM. Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychol Methods*. 2009;14(2):81–100.

3. Margolis R, Derr L, Dunn M, Huerta M, Larkin J, Sheehan J, et al. The National Institutes of Health's Big Data to Knowledge (BD2K) initiative: Capitalizing on biomedical big data. *J Am Med Inform Assoc*. 2014;21(6):957–958.

4. Nesselroade JR. The warp and woof of the developmental fabric. In: Downs R, Liben L, Palermo DS, eds. *Visions of Development, the Environment, and Aesthetics: The Legacy of Joachim F Wohlwill*. Hillsdale, NJ: Lawrence Erlbaum Associates;1991:213–240.

5. Lee WC, Ban JC. A comparison of IRT linking procedures. *Appl Meas Educ*. 2010;23(1):23–48.

6. Skubic M, Jimison H, Keller J, Popescu M, Rantz M, Kaye J, et al. A framework for harmonizing sensor data to support embedded health assessment. Paper presented at: 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2014;1747–1751.

7. Nahum-Shani I, Potter, L.N., Lam, C.Y., Yap, J., Moreno, A., Stoffel, R., et al. The mobile-assistance for regulating smoking (MARS) micro-randomized trial design protocol. 2021;110:106513.

8. Wen CK, Liao Y, Maher J, Huh J, Belcher B, Dzubur E, et al. Relationships among affective states, physical activity, and sedentary behavior in children: Moderation by perceived stress. *Health Psychol*. 2018;37:904–914.

9. Kuppens P, Allen NB, Sheeber LB. Emotional inertia and psychological adjustment. *Psychol Sci*. 2010;21:984–991.

10. Ou L, Hunter MD, Chow SM. What's for dynr: A package for linear and nonlinear dynamic modeling in R. *R Journal*. 2019; Available at https://journal.r-project.org/archive/2019/RJ-2019-012/index.html.

11. Oravecz Z, Tuerlinckx F, Vandekerckhove J. Bayesian data analysis with the bivariate hierarchical Ornstein-Uhlenbeck process model. *Multivariate Behav Res*. 2016;51(1):106–119.

12. The Neuroinformatics Research Group. *DPdash User Documentation*. Harvard University; 2018. Available at https://sites.google.com/g.harvard.edu/dpdash; https://github.com/harvard-nrg/dpdash.

13. Ponnada A, Cooper S, Tang Q, Thapa-Chhetry B, Miller JA, John D, Intille S. Signaligner Pro: A tool to explore and annotate multiday raw accelerometer data. Paper presented at: IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops) [Internet]. 2021; Available at https://crowdgames.github.io/signaligner.

14. Ponnada A, Cooper S, Thapa-Chhetry B, Miller JA, John D, Intille S. Designing videogames to Crowdsource accelerometer data

15. Thapa B, Arguello D, John D, Intille S. Detecting sleep and nonwear in 24-hour wrist accelerometer data from the National Health and Nutrition Examination Survey [online ahead of print on June 23, 2022]. *Med Sci Sports Exerc*. doi:10.1249/MSS.0000000000002973.

16. Vallée J, Cadot E, Roustit C, Parizot I, Chauvin P. The role of daily mobility in mental health inequalities: the interactive influence of activity space and neighbourhood of residence on depression. *Soc Sci Med*. 2011;73(8):1133–1144.

17. Zhou S, Li Y, Chi G, Oravecz Z, Bodovski Y, Friedman N, et al. GPS2space: An open-source python library for spatial measure extraction from GPS data. *Behav Res Methods*. 2021;1(2):127–155.

18. Li Y, Oravecz Z, Zhou S, Bodovski Y, Barnett IJ, Chi G, et al. Bayesian forecasting with a regime-switching zero-inflated multilevel Poisson regression model: An application to adolescent alcohol use with spatial covariates. *Psychometrika*. 2022;87(2):376–402

19. Lyden K, Petruski N, Mix S, Staudenmayer J, Freedson P. Direct Observation is a Valid Criterion for Estimating Physical Activity and Sedentary Behavior. *J Phys Act Health*. 2014;11:860–3.

20. John D, Tang Q, Albinali F, Intille S. An open-source monitor-independent movement summary for accelerometer data processing. *J Meas Phys Behav*. 2019;2(4):2681–2281.

21. Belcher B, Wolff-Hughes D, Dooley E, Staudenmayer J, Berrigan D, Eberhardt MS, et al. US Population-referenced percentiles for Wrist-Worn Accelerometer-derived Activity. *Med Sci Sports Exerc*. 2021;53(11):2455–2464.

22. Rabbi M, Yap J, Philyaw-Kotov ML, Klasnja P, Bonar EE, Cunningham RM, et al. Translating strategies for promoting engagement in mobile health: A proof-of-concept micro-randomized trial. *Health Psychol*. 2021; 40(12):974–987.

23. Rubin DB. Inference and missing data. *Biometrika*. 1976;63(3):581–592.

24. Beck AT, Kovacs M, Weissman A. Assessment of suicidal intention: the scale for suicide ideation. *J Consult Clin Psychol*. 1979;47(2):343–352.

25. Holi M, Pelkonen M, Karlsson L, Kiviruusu O, Ruuttu T, Heilä H, et al. Psychometric properties and clinical utility of the Scale of Suicidal Ideation (SSI) in adolescents. *BMC Psychiatry*. 2005;5:8.

26. First MB. Structured Clinical Interview for the DSM (SCID). In: *The Encyclopedia of Clinical Psychology [Internet]*. John Wiley & Sons, Ltd; 2015:1–6. Available at https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118625392.wbecp351

27. Hagströmer M, Oja P, Sjöström M. The International Physical Activity Questionnaire (IPAQ): a study of concurrent and construct validity. *Public Health Nutr*. 2006;9(6):755–762.

28. Kaphingst K, Ali J, Taylor H, Kass N. Rapid estimate of adult literacy in medicine: Feasible by telephone? *Family Med*. 2010 Jul;42:467–468.

29. Onnela JP. *Digital Phenotyping and Beiwe Research Platform* [Internet]. 2021. Available at https://www.hsph.harvard.edu/onnela-lab/beiwe-research-platform/.

30. Brick T. *Wear-IT: Real Time Science in Real Life* [Internet]. 2018. Available at https://sites.psu.edu/realtimescience/projects/wear-it/.

31. Lind M, Byrne M, Wicks G, Smidt A, Allen N. The Effortless Assessment of Risk States (E.A.R.S.) tool. *JMIR Mental Health*. 2018;5:e10334.

32. Klasnja P, Smith S, Seewald NJ, Lee A, Hall K, Luers B, et al. Efficacy of contextually tailored suggestions for physical activity: A micro-randomized optimization trial of HeartSteps. *Ann Behav Med*. 2018;53(6):573–582.

33. Hossain SM, Hnat T, Saleheen N, Nasrin NJ, Noor J, Ho BJ, et al. mCerebrum: An mHealth software platform for development and validation of digital biomarkers and interventions. In: *The ACM Conference on Embedded Networked Sensor Systems (SenSys)* [Internet]. ACM; 2017. Available at https://md2k.org/images/papers/software/mCerebrum-SenSys-2017.pdf