MOBILEWELL400+: A LARGE-SCALE MULTIVARIATE LONGITUDINAL MOBILE DATASET FOR INVESTIGATING INDIVIDUAL AND COLLECTIVE WELL-BEING

TECHNICAL REPORT

Oresti Banos*
Research Centre for Information and Communication Technologies (CITIC-UGR)
University of Granada
oresti@ugr.es

Miguel Damas
Research Centre for Information and Communication Technologies (CITIC-UGR)
University of Granada

Carmen Goicoechea
Mind, Brain and Behavior Research Center (CIMCYC)
University of Granada

Pandelis Perakakis
Department of Social, Work, and Differential Psychology
Complutense University of Madrid

Hector Pomares
Research Centre for Information and Communication Technologies (CITIC-UGR)
University of Granada

Ciro Rodriguez-Leon
Research Centre for Information and Communication Technologies (CITIC-UGR)
University of Granada

Daniel Sanabria
Mind, Brain and Behavior Research Center (CIMCYC)
University of Granada

Claudia Villalonga
Research Centre for Information and Communication Technologies (CITIC-UGR)
University of Granada

ABSTRACT

This study engaged 409 participants over a period spanning from July 10 to August 8, 2023, ensuring representation across various demographic factors: 221 females, 186 males, 2 non-binary, year of birth between 1951 and 2005, with varied annual incomes and from 15 Spanish regions. The MobileWell400+ dataset, openly accessible, encompasses a wide array of data collected via the participants' mobile phone, including demographic, emotional, social, behavioral, and well-being data. Methodologically, the project presents a promising avenue for uncovering new social, behavioral, and emotional indicators, supplementing existing literature. Notably, artificial intelligence is considered to be instrumental in analysing these data, discerning patterns, and forecasting trends, thereby advancing our comprehension of individual and population well-being. Ethical standards were upheld, with participants providing informed consent. Data collection involved smartphone app installation, passive sensor data gathering, and regular surveys, facilitating a comprehensive understanding of participants' well-being dynamics.

*Corresponding author

Keywords Smartphones · Mobile data · Human behaviour · Physical activity · Social interactions · Emotional states · Well-being

1 Introduction

The emphasis on well-being and mental health has grown stronger, underscoring the necessity for dependable techniques to observe and evaluate the complexities of daily life. Conventional methods are constrained by retrospective approaches and fixed measures, resulting in gaps in comprehension. Smartphone technology presents a solution by allowing real-time monitoring of social, behavioral, and emotional data. Through objective sensor data and digital questionnaires gathered via users' mobile devices, a thorough understanding of well-being dynamics can be achieved, aiding in the identification of patterns and the prediction of future trajectories.

To assess the validity of this approach, a study was conducted spanning nearly a month, from July 10 to August 8, 2023, involving 409 participants selected to represent the Spanish population. The MobileWell400+ dataset, available for public access, encompasses demographic information, health-related inquiries, and data on emotions, social interactions, behaviors, and overall well-being. The methodology and resulting data from this study offer the potential to reveal new indicators related to social, behavioral, and emotional trends, supplementing those found in existing literature. Notably, artificial intelligence plays a crucial role in analyzing these data, identifying patterns, and forecasting trends, thus advancing our understanding of well-being in modern society.

2 Study description

The participant sample for this study was sourced through a market research company that established a panel of individuals meeting the specific requirements of the study. The selection process employed a quota stratified sampling methodology to ensure representation of the Spanish population in terms of gender, age, location, and annual income.

Before the study began, candidates were provided with a detailed information sheet that outlined the study's parameters, including its duration and start date, the number of surveys to be completed, and the frequency of their completion, as well as data privacy considerations. The rewards for participating were also outlined, with a minimum of 80% completion rate of the daily surveys required to receive the reward.

All study procedures were conducted in accordance with relevant ethical guidelines and regulations. The study was approved by the Ethical Committee of the University of Granada under reference number 2214/CEIH/2021. Prior to participating, all individuals provided informed consent and confirmed that they were at least 18 years old. Participation in the study was strictly voluntary and all data collected was anonymous and confidential. The study adhered to the ethical standards outlined in the Declaration of Helsinki.

A total of 409 individuals completed the minimum required registration period (i.e., July 10, 2023 to August 8, 2023). The participants were composed of 221 (54.0%) females, 186 (45.5%) males, and 2 (0.5%) other, with ages ranging from 18 to 72 years (mean±std age=45.2±15.0). Annual net income was classified according to the criteria established by the Spanish Statistical Institute (INE). To ensure representation of the diverse Spanish population, the Nielsen Geographic Zones criteria were used. Of the 409 participants, 236 (57.7%) completed 70% or more of the surveys.

For the data collection, the study participants were required to install a custom mobile app (`https://github.com/orestibl/postcovid-ai/tree/main/mobile-app/v2`) on their smartphones, enter the identification number provided by the recruitment company, grant the necessary permissions for the app's proper functioning, and provide their digital informed consent to participate in the study. As a part of the enrollment process, participants were asked to complete an initial survey, which included demographic, socioeconomic, health-related, and employment information, as well as questionnaires aimed at measuring their well-being and other social aspects.

Once enrolled, the app initiated the passive data collection through the smartphone's sensors, including physical activity recognition, and indicators of social activity such as connection type, screen usage, WiFi networks, ambient light, and noise. Moreover, the self-reported emotional and social interaction data was collected using the ESMs implemented through the app. The app pushed notifications to participants at three designated times per day, randomly distributed between 10:00-11:00, 16:00-17:00, and 22:00-23:00. The notification persisted for one hour before disappearing. Upon opening the notification, the app prompted participants to complete the corresponding survey, which was then transmitted to the data storage server. In addition to the daily surveys, the app prompted participants to complete weekly questionnaires on their health, loneliness, and life satisfaction, to monitor any changes over time.

At the end of the study, participants were asked to complete a final survey, which included questionnaires aimed at measuring their well-being and other social aspects.

Table 1: Overview of the data collected in the population study.

Data type	**Variable**	**Instruments or Indicators**
Sensor	Activity recognition	Detected physical activities
	Wifi	Connections to WiFi networks
	Connectivity	Type of connections with the network
	Light	Ambient light measurements
	Noise	Ambient noise measurements
	Screen	Smartphone screen status
Initial Survey	Participant's characteristics	Questions on demographic, socioeconomic, health, and employment related data
	Well-being measures	International PANAS Short Form (I-PANAS-SF)
		General life satisfaction and seven domain of life
		Flourishing Scale (FS)
		Patient Health Questionnaire – 9 (PHQ-9)
		Generalized Anxiety Disorder Scale (GAD-7)
		Brief Resilience Scale (BRS)
		Acceptance And Action Questionnaire – II (AAQ-II)
	Social measures	Three-Item Loneliness Scale (TILS)
		Perceived social economic inequality
Daily Survey	Affect	Valence
		Energetic Arousal
		Tense Arousal
	Emotional event	Report on any remarkable situations at the emotional level
	Social interaction	Number of social interactions
		Quality of the social interactions
Weekly Survey	Health, loneliness, and life satisfaction	Questions on health
		Three-Item Loneliness Scale (TILS)
		General life satisfaction and seven domain of life
Final survey	Well-being measures	International PANAS Short Form (I-PANAS-SF)
		General life satisfaction and seven domain of life
		Flourishing Scale (FS)
		Patient Health Questionnaire – 9 (PHQ-9)
		Generalized Anxiety Disorder Scale (GAD-7)
		Brief Resilience Scale (BRS)
		Acceptance And Action Questionnaire – II (AAQ-II)
	Social measures	Three-Item Loneliness Scale (TILS)
		Perceived social economic inequality

Table 1 outlines the different data types, variables, and indicators collected via the phone.