Artificial Intelligence Assisted Follow Up for Gastrointestinal Disorders and Emotional Well Being

Mariela Gonzalez

Correspondence to: Mariela Gonzalez Email: mgonzalez4@uc.cl

Afiliación

Abstract

Introduction: Gastrointestinal disorders involve interacting digestive, emotional, sleep, and behavioral factors requiring longitudinal assessment. **Objective:** To evaluate 12-week feasibility, engagement, and clinical associations of a mobile AI-driven mental health assistant in adults with gastrointestinal disorders. **Methods:** This prospective comparative study enrolled 186 Venezuelan adults with irritable bowel syndrome, inflammatory bowel disease, functional dyspepsia, or gastroesophageal reflux disease. Participants were assigned to a mobile AI assistant plus standard medical care (n=94) or standard medical care (n=92). The intervention combined conversational support, journaling, and emotional self-reports with required smart devices monitoring (e.g., smartphone, smartwatch, and smart ring). Both groups underwent assessment of sleep, resting heart rate, nocturnal heart-rate variability, and daily steps. The primary outcome was symptomatic days per week. GAD-7 and PHQ-9 assessed anxiety and depression; adjusted differences included 95% confidence intervals. **Results:** Participant retention was 91.5% in the intervention group and 85.9% in the standard medical care group, with fewer symptomatic days, reduced anxiety and depression, and 28 additional minutes of sleep nightly. Resting heartrate decreased by 3.6 beats/min and nocturnal RMSSD increased by 5.1 ms relative to comparison; both confidence intervals excluded zero. Daily-step differences were nonsignificant. Daily check-in completion among enrolled participants was 74.8%. **Conclusions:** The mobile AI intervention was associated with improvements in gastrointestinal symptoms, emotional outcomes, sleep, and autonomic measures.

Keywords: Artificial intelligence, mental health, gastrointestinal diseases, smart devices.

Seguimiento asistido por inteligencia artificial para trastornos gastrointestinales y bienestar emocional

Resumen

Introducción: Los trastornos gastrointestinales involucran factores digestivos, emocionales, del sueño y conductuales que requieren evaluación longitudinal. **Objetivo:** Evaluar la factibilidad, participación y asociaciones clínicas de un asistente móvil de salud mental basado en IA durante 12 semanas en adultos con trastornos gastrointestinales. **Métodos:** Estudio comparativo prospectivo con 186 adultos venezolanos con síndrome de intestino irritable, enfermedad inflamatoria intestinal, dispepsia funcional o reflujo gastroesofágico. Asignados a asistente móvil de IA más atención médica habitual (n=94) o atención médica habitual (n=92). La intervención combinó apoyo conversacional, diarios, autorreportes emocionales y monitorización mediante dispositivos inteligentes (por ejemplo, teléfono, reloj y anillo). Se evaluaron sueño, frecuencia cardíaca en reposo, variabilidad nocturna de la frecuencia cardíaca y pasos diarios. El desenlace principal fueron los días sintomáticos semanales. GAD-7 y PHQ-9 midieron ansiedad y depresión. **Resultados:** La retención fue del 91,5 % y 85,9 %, respectivamente. El grupo de intervención presentó menos días sintomáticos, menor ansiedad y depresión, y 28 minutos adicionales de sueño por noche. La frecuencia cardíaca disminuyó 3,6 latidos/min y el RMSSD nocturno aumentó 5,1 ms; ambos intervalos de confianza excluyeron cero. Las diferencias en pasos diarios no fueron significativas. La cumplimentación diaria fue del 74,8 %. **Conclusiones:** La intervención se asoció con mejoras gastrointestinales, emocionales, del sueño y autonómicas.

Palabras clave: Inteligencia artificial, salud mental, enfermedades gastrointestinales, dispositivos inteligentes.

Introduction

Chronic gastrointestinal diseases create a clinical burden that extends beyond isolated digestive complaints. Abdominal pain, diarrhea, constipation, bloating, reflux, fatigue, and sleep disturbance fluctuate in daily life, whereas health systems often evaluate them episodically. In disorders of gut-brain interaction, Rome IV emphasizes that symptom generation involves bidirectional mechanisms including visceral hypersensitivity, motility, microbiota, central processing, stress, and psychosocial factors.[1-4]

Irritable bowel syndrome (IBS) illustrates this challenge. It is common, heterogeneous, and associated with impaired quality of life, work productivity loss, and repeated health-care use. Contemporary reviews emphasize education, subtype-based treatment, dietary intervention, pharmacologic options, psychological approaches, and realistic therapeutic goals.[5-10] In practice, many patients struggle to translate recommendations into sustained routines or identify individual exacerbation patterns.

Similarly, gastroesophageal reflux disease (GERD) and functional dyspepsia present a related problem: common symptoms, variable intensity, and complex relationships with meals, sleep, anxiety, body weight, and medication use.

How to cite this article: Gonzalez M. Artificial Intelligence Assisted Follow Up for Gastrointestinal Disorders and Emotional Well Being. Rev Gen. 2026;80(4):193-99. doi: 10.61155/gen.v80i4.846

© Revista Gen. Caracas, Venezuela - ISSN 2477-975X

GERD guidelines and diagnostic consensus statements stress careful history, appropriate therapeutic trials, and objective testing for refractory or alarm symptoms.[11,12] Clinical decisions may improve when patients can accurately reconstruct frequency, context, and treatment response.

In inflammatory bowel disease (IBD), current management extends beyond symptom control toward inflammatory control, mucosal healing, and disability prevention. STRIDE and STRIDE-II established treat-to-target strategies based on symptoms, biomarkers, and endoscopic assessment when appropriate.[13-18] Nevertheless, symptoms can be discordant with endoscopic or histologic activity, and some patients in objective remission continue to experience fatigue, pain, urgency, or anxiety. Patient-reported measures therefore complement but do not replace biomarkers.[15,17,18]

Digital follow-up may narrow the gap between daily patient experience and clinical encounters. Mobile devices and wearables (e.g., smartphone, smartwatch, and smart ring) can repeatedly capture symptoms, sleep, activity, and physiological signals, potentially identifying personalized patterns and supporting shared decision-making.[19-25] However, data availability alone does not ensure clinical utility; data quality, interoperability, equity, privacy, analytical validation, and medical interpretation are essential conditions for responsible use.[20-22]

Measurement quality is a central issue in digital research. High completion rates do not prove clinical benefit, and a translated questionnaire is not automatically valid in every Spanish-speaking population. Psychometric interpretation depends on evidence for reliability, factor structure, criterion validity, and measurement equivalence in the population of interest. Conversely, user engagement with digital interventions can be measured directly through prespecified event logs without converting behavior into a subjective scale. This study therefore separated clinical symptom measures, Venezuelan supported psychometric instruments, and objective engagement indicators. [26-35]

ApoyAI is a mobile application featuring an artificial intelligence assistant designed to support preventive mental well-being. It helps users explore relationships among emotions, voice or text narratives, self-reported experiences, and physiological indicators, including sleep, heart rate, heart-rate variability, and daily step count, collected through smart devices.[36,37] It was not designed as a psychiatric diagnostic tool or as a substitute for psychologists, psychiatrists, gastroenterologists, emergency services, or prescribed treatment. This study evaluated its existing capabilities in a gastrointestinal clinical population without disease-specific software adaptation. Its relevance to gastroenterology lies in the opportunity to examine the gut-brain, sleep, behavioral, and autonomic context associated with fluctuations in digestive symptoms.[1-10,19-25]

Study objective

To evaluate the feasibility, objective engagement, and clinical associations of twelve-week use of ApoyAI, an artificial-intelligence mental-health and well-being assistant, among Venezuelan adults with IBS, IBD, functional dyspepsia, or GERD, using psychometric scales supported by Venezuelan evidence and a contemporaneous standard medical comparison group.

Material and Methods

Design and population

We designed a pragmatic, prospective, comparative, parallel-group study with 12 weeks of follow-up. All patients were evaluated at Clínica El Ávila, Caracas, Venezuela, between January and July 2026. Eligible participants were adults aged 18 years or older with a gastroenterologist-confirmed diagnosis of IBS, IBD, functional dyspepsia, or GERD; access to a compatible smartphone; and willingness to complete follow-up assessments.

The intervention cohort used ApoyAI with required smart device monitoring (e.g., smartphone, smartwatch, and smart ring) plus standard medical care. Comparators received standard medical without platform access; identical passive biometric collection in both groups supported outcome comparison, without feedback to comparators.

The twelve-week duration was selected to capture persistence of use after initial novelty, permit repeated observation of fluctuating symptoms, and evaluate whether engagement was sustained beyond three weeks. Diagnoses were established from clinical records using contemporary specialty criteria: Rome IV for disorders of gut-brain interaction, guideline-supported evaluation for GERD, and established clinical, endoscopic, histologic, and biomarker evidence for IBD when applicable.[1-8,11-18]

Exclusion criteria were uninvestigated alarm symptoms, major gastrointestinal surgery during the previous three months, pregnancy, inability to complete Spanish-language digital questionnaires, hospitalization at enrollment, severe uncontrolled psychiatric illness that precluded consent or study participation, or absence of a confirmed gastrointestinal diagnosis. The analysis included all enrolled participants with a baseline assessment and at least one post-baseline observation; missing longitudinal observations were handled in the prespecified mixed-model analysis (**Table 1**).

Table 1. Baseline demographic and clinical characteristics of participants by study group

Characteristic	Pre-release ApoyAI + standard medical care (n=94)	Standard medical (n=92)	Total (n=186)
Age, mean ± SD	42.6 ± 13.8	43.9 ± 14.1	43.2 ± 13.9
Female sex, n (%)	57 (60.6)	55 (59.8)	112 (60.2)
IBS, n (%)	35 (37.2)	33 (35.9)	68 (36.6)
IBD, n (%)	26 (27.7)	26 (28.3)	52 (28.0)
Functional dyspepsia, n (%)	17 (18.1)	17 (18.5)	34 (18.3)
GERD, n (%)	16 (17.0)	16 (17.4)	32 (17.2)
Symptomatic days/week, mean ± SD	6.5 ± 1.4	6.4 ± 1.5	6.4 ± 1.5
Previous health-app use, n (%)	31 (33.0)	29 (31.5)	60 (32.3)

ApoyAI intervention and study version

ApoyAI is a mobile application featuring an artificial intelligence assistant designed to support preventive mental well-being between conventional healthcare visits. The application helps users document emotional experiences, recognize changes in daily functioning, and explore possible relationships among emotional state, perceived stress, sleep, daily step count, physiological signals, and health-related symptoms. Its purpose is to organize information collected across different moments and sources into a longitudinal record that can support self-awareness and more informed communication with healthcare professionals.

The mental-health assistant combines conversational interaction with structured monitoring. Participants could interact with the application through voice or text, maintain personal journals, record emotions and perceived stress, and complete brief psychological assessments. The conversational component was designed to encourage users to describe how they felt, identify relevant events, and reflect on changes in mood, energy, sleep, behavior, and physical symptoms. Rather than providing unrestricted psychological counseling, the assistant generated concise supportive responses and personalized summaries based on information recorded or entered in writing by participants.

When authorized by the participant, ApoyAI could incorporate information from compatible smart devices and health-data repositories (included Apple Health or HealthKit, Android Health Connect). These data included sleep duration and patterns, resting and active heart rate, heart-rate variability, daily step count. The application organized self-reported and biometric information chronologically to identify deviations from the participant's recent individual pattern. Its AI-generated summaries and insights were intended to highlight potentially relevant observations (for example, the temporal coexistence of poorer sleep, increased perceived stress, reduced activity, emotional changes, and worsening physical symptoms) without presenting these relationships as clinically proven causes.

For the present study, the pre-release version of the ApoyAI mobile application was used with its existing mental-health and biometric monitoring capabilities in adults with gastrointestinal disorders. Participants used the application's available features to document their emotional experiences, perceived stress, daily activities, sleep, and other information supported by the application. These data were examined in the context of participants' gastrointestinal conditions to explore potential associations between mental well-being, behavioral patterns, physiological indicators, and gastrointestinal health.

ApoyAI operated as a preventive and complementary support tool. It did not independently diagnose anxiety, depression, irritable bowel syndrome, inflammatory bowel disease, gastroesophageal reflux disease, or any other psychiatric or gastrointestinal condition. It did not determine inflammatory activity, establish causal relationships, or autonomously recommend starting, discontinuing, or modifying treatment. Psychological questionnaire scores and AI-generated observations were treated as screening and monitoring information. All diagnostic interpretations, safety assessments, referrals, and therapeutic decisions remained under the responsibility of qualified healthcare professionals.

Version control and deployment context

Enrollment and follow-up occurred before the public commercial launch. Each intervention participant received a

controlled pre-release mobile build linked to the study environment, together with standardized installation, consent, onboarding, and troubleshooting instructions. This approach allowed the investigators to hold the principal study functions stable during follow-up and to distinguish research use from later production updates.

At the time of this final manuscript revision, a subsequent production version of ApoyAI is publicly available for iPhone and Android.[36,37] The public application retains the preventive mental-well-being orientation and multimodal capabilities, including voice interaction, emotional records, longitudinal insights, and connections to compatible biometric ecosystems. Public availability occurred after the study period and is reported only to clarify translational maturity. The version available in app stores incorporates changes to the interface, connectivity, data analytics, privacy, and performance; therefore, it is not technically identical to the pre-release version used in the study.

Biometric measures and gastrointestinal relevance

With participant authorization, the study build retrieved four principal biometric domains from compatible smart devices or health repositories: sleep, heart rate, heart-rate variability, and daily step count. These variables were selected because gastrointestinal disorders can be influenced by sleep disruption, autonomic arousal, stress physiology, fatigue, activity patterns, and daily behavior (**Table 2**).

Their function in this study was contextual: they helped determine whether changes in gastrointestinal symptoms occurred alongside changes in recovery or behavior. No biometric value was used independently to diagnose a gastrointestinal or psychiatric disorder, determine inflammatory activity, or trigger an autonomous treatment recommendation.[1-10,17-22]

Table 2. Biometric domains, gastrointestinal relevance, and interpretation in the study

Biometric domain	Potential relevance in gastrointestinal disorders	Interpretation in the study
Sleep duration and continuity	Sleep disruption may accompany greater pain sensitivity, nocturnal reflux, fatigue, and worsening symptom perception.	Used to contextualize weekly symptom trajectories; not a sleep-disorder or gastrointestinal diagnostic measure.
Resting and active heart rate	Provides context on autonomic arousal, exertion, intercurrent illness, medication effects, and periods of perceived stress.	Interpreted with activity and self-report because heart rate is nonspecific and can change for many non-gastrointestinal reasons.
Heart-rate variability	May reflect short-term changes in autonomic modulation and recovery relevant to stress-responsive gut-brain processes.	Analyzed relative to the participant's own recent pattern; not treated as a marker of inflammation or psychiatric diagnosis.
Daily step count	Provides behavioral context for bowel routine, functional limitation, fatigue, deconditioning, and general well-being.	Used as a contextual exposure and outcome-supporting measure; affected by work, environment, illness, and device wear.

Biometrics contextualized symptom trajectories; resting heart rate, nocturnal root mean square of successive differences (RMSSD), and daily step count were exploratory outcomes. RMSSD is a time-domain measure of heart rate variability calculated as the square root of the mean of the squared differences between successive normal-to-normal heartbeat intervals; it reflects short-term cardiac autonomic modulation, particularly parasympathetic influences.[38]

None of these measures independently diagnosed disease, established inflammatory activity, or triggered treatment. Prior research in irritable bowel syndrome (IBS) and other functional gastrointestinal disorders supports investigating heart rate variability as a complementary measure of autonomic function, alongside daily step count as an indicator of physical activity.[39,40]

Baseline and week-12 biometric values were calculated as 7-day averages, requiring at least four valid days or nights for each measure. Resting heart rate (beats/min) was obtained from device-defined resting periods. Nocturnal heart rate variability (HRV) was assessed using the root mean square of successive differences between normal-to-normal intervals (RMSSD, ms), a measure of short-term variation in the time between consecutive heartbeats. Records reporting only the standard deviation of normal-to-normal intervals (SDNN), a different HRV measure, were excluded to ensure consistency across measurements.

Daily step count required at least 16 hours of device wear per day, while sleep and RMSSD analyses required valid overnight recordings. Devices remained unchanged within each participant, and comparable devices were used in both groups.

© Revista Gen. Caracas, Venezuela - ISSN 2477-975X

https://doi.org/10.61155/gen.v80i4.846

Missing or incompatible recordings were excluded from the corresponding analysis; missing step-count data were not interpreted as zero activity.

Outcomes, instruments, and measurement procedures

The primary symptom endpoint was the mean number of symptomatic days per 7-day week. Each calendar day was classified as symptomatic when the participant reported at least one moderate-or-worse gastrointestinal symptom (abdominal pain, bloating/distension, diarrhea, constipation, reflux/regurgitation, nausea, or postprandial distress) or gastrointestinal interference with usual activities. Daily status was collected in ApoyAI for intervention participants and through an equivalent weekly study form for comparators. Weekly values ranged from 0 to 7 days; lower values indicated fewer symptomatic days. This count endpoint was deliberately treated as an operational diary outcome rather than a validated psychometric scale.

Anxiety symptoms were measured with the seven-item Generalized Anxiety Disorder scale (GAD-7). Each item asks how often a symptom occurred during the preceding two weeks and is scored from 0 (not at all) to 3 (nearly every day), producing a 0-21 total; conventional severity bands are 0-4 minimal, 5-9 mild, 10-14 moderate, and 15-21 severe.[29] Depressive symptoms were measured with the nine-item Patient Health Questionnaire (PHQ-9), using the same 0-3 response structure and a 0-27 total; conventional bands are 0-4 minimal, 5-9 mild, 10-14 moderate, 15-19 moderately severe, and 20-27 severe.[30] Both were administered in Spanish at baseline, week 8, and week 12. A questionnaire was scored only when all items were completed. Item 9 of the PHQ-9 triggered the study's predefined clinical safety review; neither scale was used alone to establish a psychiatric diagnosis.

Instrument selection was based on Venezuelan evidence. Borges Grün[31] evaluated adaptation and initial validity of the Patient Health Questionnaire in Venezuelan primary care, including reliability and convergent-validity analyses, receiver-operating-characteristic analysis against the MINI Plus, and application of the PHQ-9 and GAD-7 in 814 adults from a Caracas primary-care center. Additional GAD-7 evidence comes from Venezuelan migrant samples with reliability, factor-structure, and measurement-equivalence analyses,[32] and from 227 medical students at a Venezuelan public university.[33] Complementary PHQ-9 evidence includes a Spanish-speaking immigrant study with a substantial Venezuelan-born subgroup.[34] These studies support use in Venezuelan populations but do not eliminate the need for diagnosis-specific validation in gastroenterology; scores were therefore interpreted conservatively as symptom-severity outcomes.

Perceived usability was assessed at week 12 with the 10-item Spanish System Usability Scale (SUS). Items were rated from 1 to 5; odd-item responses contributed response minus 1, even-item responses contributed 5 minus responses, and the summed score was multiplied by 2.5 to produce a 0-100 total, with higher values indicating greater perceived usability. Because no Venezuelan psychometric validation was identified, SUS was analyzed descriptively rather than as an inferential endpoint.[35]

Objective engagement was obtained from auditable application logs. Week-12 retention meant completion of the final study assessment. An active day required a participant-initiated event: completion of a symptom check-in, entry of a contextual exposure, or opening of a weekly report; background synchronization alone did not qualify. Check-in completion was the number of completed daily check-ins divided by 84 scheduled days. Weekly-report exposure counted the first verified opening of each of 12 reports. Among connected-device users, a valid biometric day required the domain-specific wear or overnight criteria above; coverage was valid days divided by days from successful connection through week 12.

Statistical analysis

Statistical analyses were performed using R software. Continuous variables were summarized using means and standard deviations or medians and interquartile ranges; categorical variables were reported as frequencies and percentages. Baseline characteristics were compared descriptively between groups.

Changes in outcomes over 12 weeks were analyzed using linear mixed-effects models accounting for repeated measurements within participants. Adjusted between-group differences in change from baseline to week 12 were reported with 95% confidence intervals (CIs). GAD-7 and PHQ-9 scores were analyzed as continuous outcomes. Retention was compared using risk ratios and 95% CIs.

Analyses included participants with baseline data and at least one follow-up measurement. Mixed- effects models used all available data under a missing-at-random assumption, with a complete-case sensitivity analysis.

Diagnosis-specific and biometric comparisons were exploratory, without multiplicity adjustment. Statistical significance was defined as two-sided $p<0.05$; nominal p values, effect sizes, and confidence intervals guided interpretation.

Ethical considerations

Patients were evaluated at La Clínica El Ávila, Caracas, Venezuela, from January to July 2026. The study followed the Declaration of Helsinki, with digital informed consent obtained before study procedures. Participation was voluntary, and identifiable data were accessible only to authorized personnel.

ApoyAI safeguards sensitive health data in compliance with GDPR, HIPAA, Latin American privacy laws, including Brazil's LGPD and Chile's Law 19,628, ISO/IEC 27001 and ISO/IEC 27701.

Results

Retention, use, and usability

Among 213 screened patients, 186 were enrolled and contributed a baseline measurement: 94 received the controlled pre-release ApoyAI build plus standard medical care and 92 received standard medical care alone.

By week 12, 86 intervention participants (91.5%) and 79 comparators (85.9%) completed the final assessment. Intervention participants recorded a median of 68 participant-initiated active days (IQR 59-76) of 84, completed 74.8% of scheduled daily check-ins, and opened a mean of 9.1±2.3 of 12 weekly reports. Among participants with a successfully connected device, 78.4% of eligible days met the prespecified biometric-validity rule (**Table 3**).

Table 3. Participant Retention, App Engagement, and Usability in the ApoyAI Group

Objective engagement indicator	ApoyAI result
Week-12 retention	86/94 (91.5%)
Participant-initiated active days	68 (IQR 59-76) of 84 days
Scheduled daily check-ins completed	74.8%
Valid biometric coverage among connected users	78.4% of eligible days
Weekly reports opened	9.1 ± 2.3 of 12 reports

Note: Values are presented as n/N (%), median (interquartile range), or mean ± standard deviation, as appropriate. Biometric coverage was calculated among participants with successfully connected devices. SUS scores were interpreted descriptively.

These engagement indicators were generated from timestamped event logs and therefore represent observed platform behavior rather than perceived ease of use. Background synchronization was excluded from active day counts to reduce inflation by passive data transfer. The mean week-12 SUS score was 83.9/100 and was treated as a descriptive indicator of perceived usability because Venezuelan psychometric validation was unavailable.

Clinical and well-being outcomes

Mean days with moderate or severe gastrointestinal symptoms decreased from 6.5±1.4 to 4.1±1.7 days/week in the intervention group and from 6.4±1.5 to 5.2±1.8 with standard medical care. The adjusted difference was −1.1 days/week (95% CI −1.6 to −0.6; p<0.05), equivalent to approximately one additional symptom-free day per week with digital follow-up (**Table 4**).

GAD-7 scores decreased from 9.2±4.8 to 6.1±4.2 in the intervention group and from 9.0±4.9 to 7.8±4.6 in comparators (adjusted difference −1.8; 95% CI −3.0 to −0.6). PHQ-9 scores decreased from 8.7±5.1 to 5.9±4.6 and from 8.5±5.0 to 7.2±4.9, respectively (adjusted difference −1.4; 95% CI −2.6 to −0.2). Both differences were significant (p<0.05).

These scores reflected changes in symptom severity, not psychiatric diagnoses. Nightly sleep increased by 0.7 hours in the intervention group and 0.2 hours in comparators (adjusted difference +28 minutes; 95% CI +11 to +45; p<0.05). In exploratory analyses, adjusted resting heart rate decreased by 3.6 beats/min (95% CI −5.3 to −1.9; nominal p<0.001), and nocturnal RMSSD increased by 5.1 ms (95% CI 1.9 to 8.3; nominal p=0.002). Daily steps increased by an adjusted 320 steps/day (95% CI −210 to +850; nominal p=0.24), without a significant between-group difference (**Table 4**).

Diagnosis-specific analysis

The direction of symptom-day change was consistent across the four diagnostic strata. IBS showed the largest mean reduction, particularly in days marked by pain, bloating, altered bowel habits, or functional interference. In functional dyspepsia, the diary was most informative for postprandial symptoms and meal-related patterns. In GERD, nocturnal symptoms, medication timing, and sleep context were prominent.

In IBD, symptom monitoring improved longitudinal visibility but was explicitly interpreted alongside biomarkers and endoscopic information when clinically indicated; digital symptom improvement was never treated as proof of inflammatory control (**Table 5**).

Table 4. Changes in gastrointestinal symptoms, emotional well-being, and sleep from baseline to week 12

Variable	ApoyAI baseline	ApoyAI week 12	Comparator baseline	Comparator week 12	Adjusted difference (95% CI)
Symptomatic days/week	6.5 ± 1.4	4.1 ± 1.7	6.4 ± 1.5	5.2 ± 1.8	-1.1 (-1.6 to -0.6)*
GAD-7 (0-21)	9.2 ± 4.8	6.1 ± 4.2	9.0 ± 4.9	7.8 ± 4.6	-1.8 (-3.0 to -0.6)*
PHQ-9 (0-27)	8.7 ± 5.1	5.9 ± 4.6	8.5 ± 5.0	7.2 ± 4.9	-1.4 (-2.6 to -0.2)*
Nightly sleep	6.1 ± 1.2 h	6.8 ± 1.1 h	6.2 ± 1.1 h	6.4 ± 1.2 h	+28 min (+11 to +45)*
Resting heart rate (beats/min)	74.8 ± 9.6	70.1 ± 9.1	74.2 ± 9.4	73.1 ± 9.2	−3.6 (−5.3 to −1.9)*
Nocturnal HRV RMSSD (ms)	30.4 ± 12.1	36.6 ± 13.4	31.0 ± 12.5	32.1 ± 12.8	+5.1 (+1.9 to +8.3)*
Daily steps (steps/day)	5,420 ± 2,140	6,080 ± 2,280	5,510 ± 2,090	5,840 ± 2,160	+320 (−210 to +850)

* p<0.05
Values are mean change ± SD. Negative values indicate a reduction in symptomatic days per week from baseline to week 12. Biometric outcomes were exploratory. p values for these outcomes are nominal and were not adjusted for multiple comparisons.

Table 5. Diagnosis-specific changes in gastrointestinal symptomatic days

Diagnosis	n ApoyAI	Change in symptomatic days/week: ApoyAI	Change: comparator	Clinical interpretation
IBS	35	-2.7 ± 1.8	-1.3 ± 1.6	Largest reduction; pain, bloating, bowel pattern, and interference predominated
IBD	26	-2.0 ± 1.6	-1.0 ± 1.5	Symptoms improved; objective inflammatory assessment remained necessary
Functional dyspepsia	17	-2.3 ± 1.7	-1.1 ± 1.6	Postprandial and meal-related patterns were prominent
GERD	16	-2.1 ± 1.5	-1.2 ± 1.5	Timing, adherence, posture, and nocturnal symptoms were emphasized

Values are mean change ± SD. Negative values indicate a reduction in symptomatic days per week from baseline to week 12.

Discussion

This 12-week comparative study suggests that a pre-release version of a mobile application featuring an artificial intelligence assistant for preventive mental well-being can support longitudinal follow-up in adults with gastrointestinal disorders. Application use was associated with sustained engagement, approximately one additional symptom-free day per week relative to standard medical care, modest reductions in anxiety and depressive symptom scores, and longer nightly sleep. These findings apply to the specific research version and study workflow evaluated and should not be generalized to subsequent releases without further assessment.

ApoyAI combines emotional self-monitoring with information from different sources, including voice or text narratives and biometric indicators. This approach may help explore relationships between mental well-being, sleep, physiological measures, and gastrointestinal symptoms over time.

These relationships are particularly relevant to disorders of gut-brain interaction, in which symptoms may be associated with emotional distress, autonomic regulation, behavior, sleep, and prior symptom expectations. [1-8,17,18] Importantly, the study evaluated ApoyAI's existing capabilities in a gastrointestinal population; no disease-specific module or software adaptation was developed.

The study was not designed to isolate individual component effects. Therefore, the respective contributions of conversational support, emotional self-monitoring, weekly reports, and biometric feedback cannot be determined. Future component-based studies could establish which features contribute most to observed outcomes.

The transition of ApoyAI from a pre-release research build to public availability through the Apple App Store and Google Play Store underscores the importance of version control. Reproducibility requires documentation of the exact study build, enabled features, data pipeline, device compatibility, and subsequent software changes. Public availability demonstrates distribution beyond the research setting but

© Revista Gen. Caracas, Venezuela - ISSN 2477-975X

https://doi.org/10.61155/gen.v80i4.846

does not establish clinical effectiveness, regulatory status, or technical equivalence between research and production versions.[26-28,36,37]

The gastrointestinal symptom findings are clinically plausible. IBS and functional dyspepsia involve heterogeneous symptom patterns, visceral sensitivity, dietary exposures, sleep disturbance, and psychological comorbidity.[1-8,17,18] Daily diaries may reduce recall error and help identify patterns that are difficult to recognize during episodic consultations. However, the symptomatic-day measure used here was an operational endpoint, not a validated gastrointestinal scale. Although readily interpretable, it weights qualifying symptoms equally and does not replace diagnosis-specific instruments.

For GERD, longitudinal records may help document symptom timing, medication adherence, meals, body position, and sleep. Their potential value lies in improving clinical history and supporting targeted evaluation, rather than confirming the diagnosis or directly modifying acid exposure. Current guidance emphasizes careful symptom phenotyping and cautions against assuming that persistent upper gastrointestinal complaints are always reflux-mediated.[9,10]

In IBD, symptom improvement must be distinguished from inflammatory control. Symptoms may improve despite persistent inflammation or continue during endoscopic remission.[11-16] Digital symptom tracking may complement follow-up by making changes visible between visits, but reports must be interpreted alongside fecal calprotectin, C-reactive protein, imaging, endoscopy, and clinical assessment when appropriate. Symptom improvement alone cannot establish disease control.

Engagement remained substantial over 12 weeks: participants initiated activity on a median of 68 of 84 days, completed approximately three-quarters of scheduled check-ins, and opened most weekly reports. Prespecified measures derived from timestamped application logs provided auditable numerators and denominators.

The GAD-7 and PHQ-9 are brief, widely used measures, but their relevance depends on evidence beyond Spanish-language availability. Venezuelan studies have reported validity evidence for PHQ measures in primary care and have examined the GAD-7 in Venezuelan migrants and medical students.[31-34] These findings support cautious assessment of symptom severity but do not establish measurement equivalence in all Venezuelan populations, particularly gastroenterology outpatients.

The adjusted between-group differences of 1.8 points on the GAD-7 and 1.4 points on the PHQ-9 indicate modest differences in reported symptom severity; they do not demonstrate treatment of anxiety or depressive disorders. Digital administration preserved item wording and response options, but these safeguards do not replace formal assessment of equivalence between digital and other administration modes. Future research should examine responsiveness, minimally important change, and longitudinal measurement invariance in Venezuelan patients with gastrointestinal disorders. Positive screening results also require appropriate clinical assessment and safety procedures.

The mean System Usability Scale (SUS) score was 83.9/100, consistent with favorable perceived usability under conventional interpretation.[35] However, the Spanish version had not been validated in a Venezuelan population, so this score should be considered exploratory. It complements, but does not replace, objective engagement measures or culturally appropriate usability evaluation.

Exploratory analyses showed lower resting heart rate and higher nocturnal RMSSD alongside longer sleep duration in the intervention group. These findings are compatible with changes in sleep-related and autonomic measures but do not establish vagal recovery, causality, or mediation of gastrointestinal improvement. A small trial of cognitive behavioral therapy (CBT) in young women with constipation-predominant IBS examined HRV and symptoms, although its population and intervention differed from those in this study.[39] Device heterogeneity, respiration, medication, fitness, and illness may also influence biometric measures.

Daily step count did not differ significantly between groups. Counts increased from 5,420 ± 2,140 to 6,080 ± 2,280 in the ApoyAI group and from 5,510 ± 2,090 to 5,840 ± 2,160 in the comparison group. The adjusted difference was +320 steps/day (95% CI −210 to +850; nominal p = 0.24). This result neither establishes equivalence nor excludes a potential benefit. Step count also fails to capture exercise intensity, which should be considered in future studies.[40]

Clinical usefulness will depend on how digital information is presented and integrated into care. Reports should prioritize symptom trajectories, medication use, alarm features, sleep disruption, and discrepancies between reported symptoms and objective markers of inflammation in IBD. Implementation should minimize clinician burden and false reassurance, with human review, clear escalation procedures, and explicit communication that the application does not diagnose or prescribe.

Several limitations warrant consideration. Nonrandomized allocation permits residual confounding and self-selection. Recruitment at a single private outpatient site in Caracas limits generalizability, while diagnostic heterogeneity reduces precision for individual disorders. The symptomatic-day endpoint has not been formally validated in a Venezuelan population. Although the GAD-7 and PHQ-9 have supporting evidence from Venezuelan populations, they have not been specifically validated in Venezuelan gastroenterology outpatients. Device ownership and

connectivity may also introduce digital inequity. The unvalidated local SUS version limits cross-cultural interpretation, and application logs measure behavior rather than satisfaction or interface quality.

Despite these limitations, the study provides a framework for evaluating AI-supported mental well-being applications in digital gastroenterology research. Its strengths include 12-week follow-up, a concurrent comparator group, an operationally defined symptom outcome, objective engagement measures, psychometric instruments supported by evidence from Venezuelan populations, and explicit attention to software versioning. The findings warrant confirmation in a randomized, adequately powered, multicenter Venezuelan trial with diagnosis-specific strata, validated gastrointestinal outcomes, prespecified missing-data sensitivity analyses, and independent verification of digital endpoints. Subsequent component-based studies could determine the contributions of conversational support, emotional self-monitoring, biometric integration, and clinical reporting.

Conclusion

Among Venezuelan adults with chronic gastrointestinal disorders, twelve-week ApoyAI plus standard medical care was associated with sustained engagement and improvements in symptomatic days, anxiety, depression, and sleep, alongside lower resting heart rate and higher nocturnal RMSSD. Daily steps did not differ significantly between groups. These exploratory biometric findings support further investigation but do not establish causality, autonomic normalization, or independent diagnostic utility.

ApoyAI contributed emotional self-reports, voice or text narratives, sleep, heart rate, heart-rate variability, and activity from the required smart devices; gastrointestinal study outcomes were examined alongside these existing sources of information. The biometric signals were most useful when interpreted as longitudinal context around each patient's gastrointestinal condition, rather than as stand-alone diagnostic endpoints.

The findings are especially relevant to IBS and functional dyspepsia, where gut-brain mechanisms and contextual symptom variation are central, but the same digital model may also assist GERD follow-up and symptom surveillance in IBD. In IBD, however, digital symptom improvement cannot establish inflammatory remission and must remain subordinate to biomarkers, imaging, endoscopy, and specialist assessment when indicated. Likewise, changes in GAD-7 and PHQ-9 represent changes in reported symptom burden, not psychiatric diagnoses or evidence that the application treated a mental disorder.

ApoyAI is publicly available through the Apple App Store and Google Play Store.[36,37] providing an accessible reference for the production application. This progression increases the practical relevance of the work while making software-version transparency essential: the production application should not be assumed to reproduce the study intervention unless its functions, algorithms, data sources, and user experience are shown to be equivalent. Public distribution also does not replace clinical validation, post-deployment safety monitoring, privacy oversight, or evaluation of digital-access inequities.

The next research step should be a multicenter randomized trial that identifies the exact software build, stratifies by gastrointestinal diagnosis, uses validated disease-specific outcomes, and prospectively examines whether engagement mediates clinical change. If confirmed, a mental-health-first multimodal assistant could become a useful adjunct to gastroenterology care by converting fragmented daily experiences into structured information for patients and clinicians, while preserving professional judgment and established diagnostic and therapeutic pathways.

Conflicts of Interest

The author declares no conflicts of interest.

Funding Source

This work was supported by the National Agency for Research and Development (ANID), Chile, through the Startup Ciencia program (SUC250204).

Este es un artículo de acceso abierto.

Fecha de recepción: 12/08/2026
Fecha de revisión: 30/08/2026
Fecha de aprobación: 29/09/2026

References

1. Drossman DA, Hasler WL. Rome IV-functional GI disorders: disorders of gut-brain interaction. Gastroenterology. 2016;150(6):1257-1261. doi:10.1053/j.gastro.2016.03.035.

2. Drossman DA. Functional gastrointestinal disorders: history, pathophysiology, clinical features, and Rome IV. Gastroenterology. 2016;150(6):1262-1279.e2. doi:10.1053/j.gastro.2016.02.032.

3. Mearin F, Lacy BE, Chang L, Chey WD, Lembo AJ, Simren M, et al. Bowel disorders. Gastroenterology. 2016;150(6):1393-1407.e5. doi:10.1053/j.gas tro.201 6.02.031.

4. Stanghellini V, Chan FKL, Hasler WL, Malagelada JR, Suzuki H, Tack J, et al. Gastroduodenal disorders. Gastroenterology. 2016;150(6):1380-1392. doi:10.1053/j.gastro.2016.02.011.

5. Ford AC, Sperber AD, Corsetti M, Camilleri M. Irritable bowel syndrome. Lancet. 2020;396(10263):1675-1688. doi:10.1016/S0140-6736(20)31548-8.

6. Chey WD, Kurlander J, Eswaran S. Irritable bowel syndrome: a clinical review. JAMA. 2015;313(9):949-958. doi:10.1001/jama.2015.0954.

7. Enck P, Aziz Q, Barbara G, Farmer AD, Fukudo S, Mayer EA, et al. Irritable bowel syndrome. Nat Rev Dis Primers. 2016;2:16014. doi:10.1038/nrdp.2016.14.

8. Camilleri M. Diagnosis and treatment of irritable bowel syndrome: a review. JAMA. 2021;325(9):865-877. doi:10.1001/jama.2020.22532.

9. Katz PO, Dunbar KB, Schnoll-Sussman FH, Greer KB, Yadlapati R, Spechler SJ. ACG clinical guideline for the diagnosis and management of gastroesophageal reflux disease. Am J Gastroenterol. 2022;117(1):27-56. doi:10.14309/ajg.0000000000001538.

10. Gyawali CP, Kahrilas PJ, Savarino E, Zerbib F, Mion F, Smout AJPM, et al. Modern diagnosis of GERD: the Lyon Consensus. Gut. 2018;67(7):1351-1362. doi:10.1136/gutjnl-2017-314722.

11. Torres J, Mehandru S, Colombel JF, Peyrin-Biroulet L. Crohn's disease. Lancet. 2017;389(10080):1741-1755. doi:10.1016/S0140-6736(16)31711-1.

12. Ungaro R, Mehandru S, Allen PB, Peyrin-Biroulet L, Colombel JF. Ulcerative colitis. Lancet. 2017;389(10080):1756-1770. doi:10.1016/S0140-6736(16)32126-2.

13. Turner D, Ricciuto A, Lewis A, D'Amico F, Dhaliwal J, Griffiths AM, et al. STRIDE-II: an update on the Selecting Therapeutic Targets in Inflammatory Bowel Disease initiative. Gastroenterology. 2021;160(5):1570-1583. doi:10.1053/j.gastro.2020.12.031.

14. Peyrin-Biroulet L, Sandborn W, Sands BE, Reinisch W, Bemelman W, Bryant RV, et al. Selecting Therapeutic Targets in Inflammatory Bowel Disease (STRIDE): determining therapeutic goals for treat-to-target. Am J Gastroenterol. 2015;110(9):1324-1338. doi:10.1038/ajg.2015.233.

15. Colombel JF, Keir ME, Scherl A, Zhao R, de Hertogh G, Faubion WA, et al. Discrepancies between patient-reported outcomes, and endoscopic and histological appearance in ulcerative colitis. Gut. 2017;66(12):2063-2068. doi:10.1136/gutjnl-2016-312307.

16. Lewis JD. The utility of biomarkers in the diagnosis and therapy of inflammatory bowel disease. Gastroenterology. 2011;140(6):1817-1826.e2. doi:10.1053/j.gastro.2010.11.058.

17. Gralnek IM, Hays RD, Kilbourne A, Naliboff B, Mayer EA. The impact of irritable bowel syndrome on health-related quality of life. Gastroenterology. 2000;119(3):654-660. doi:10.1053/gast.2000.16484.

18. Whitehead WE, Palsson O, Jones KR. Systematic review of the comorbidity of irritable bowel syndrome with other disorders. Gastroenterology. 2002;122(4):1140-1156. doi:10.1053/gast.2002.32392.

19. Hirten RP, Danieletto M, Sanchez-Mayor M, Whang JK, Lee KW, Landell K, et al. Physiological data collected from wearable devices identify and predict inflammatory bowel disease flares. Gastroenterology. 2025;168(5):939-951.e5. doi:10.1053/j.gastro.2024.12.024.

20. Canali S, Schiaffonati V, Aliverti A. Challenges and recommendations for wearable devices in digital health: data quality, interoperability, health equity, fairness. PLOS Digit Health. 2022;1(10):e0000104. doi:10.1371/journal.pdig.0000104.

21. 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. NPJ Digit Med. 2020;3:55. doi:10.1038/s41746-020-0260-4.

22. Coravos A, Khozin S, Mandl KD. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. NPJ Digit Med. 2019;2:14. doi:10.1038/s41746-019-0090-4.

23. Sim I. Mobile devices and health. N Engl J Med. 2019;381(10):956-968. doi:10.1056/NEJMra1806949.

24. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med. 2015;7(283):283rv3. doi:10.1126/scitranslmed.aaa3487.

25. Dorsey ER, Topol EJ. State of telehealth. N Engl J Med. 2016;375(2):154-161. doi:10.1056/NEJMra1601705.

26. Eysenbach G. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res. 2011;13(4):e126. doi:10.2196/jmir.1923.

27. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK; SPIRIT-AI and CONSORT-AI Working Group. Reporting

guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med. 2020;26(9):1364-1374. doi:10.1038/s41591-020-1034-x.

28. Cruz Rivera S, Liu X, Chan AW, Denniston AK, Calvert MJ; SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nat Med. 2020;26(9):1351-1363. doi:10.1038/s41591-020-1037-7.

29. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med. 2006;166(10):1092-1097. doi:10.1001/archinte.166.10.1092.

30. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16(9):606-613. doi:10.1046/j.1525-1497.2001.016009606.x.

31. Borges Grün T. Estudio sobre depresión y ansiedad mediante el PHQ en Venezuela [dissertation]. Madrid: Universidad Complutense de Madrid; 2017.

32. Sahbaz S, Cox RB Jr, Montero-Zamora P, Salas-Wright CP, Maldonado-Molina MM, Bates MM, et al. Measuring anxiety among Latino immigrant populations: within-country and between-country comparisons. Assessment. 2024;31(7):1414-1429. doi:10.1177/10731911231223715.

33. Duran P, Ortega Á, Galban N, Vera I, Díaz A, Navarro C, et al. Psychometric behaviour of the GAD-7 in medical students: structural stability, measurement equivalence and contextual sensitivity. Healthcare (Basel). 2026;14(5):563. doi:10.3390/healthcare14050563.

34. Errazuriz A, Beltrán R, Torres R, Passi-Solar A. The validity and reliability of the PHQ-9 and PHQ-2 on screening for major depression in Spanish speaking immigrants in Chile: a cross-sectional study. Int J Environ Res Public Health. 2022;19(21):13975. doi:10.3390/ijerph192113975.

35. Sevilla-Gonzalez MDR, Moreno Loaeza L, Lazaro-Carrera LS, Bourguet Ramirez B, Vázquez Rodríguez A, Peralta-Pedrero ML, et al. Spanish version of the System Usability Scale for the assessment of electronic tools: development and validation. JMIR Hum Factors. 2020;7(4):e21161. doi:10.2196/21161.

36. Digital Helth Technology Limitada. Apoyai [mobile application listing]. App Store [Internet]. Available from: https://apps.apple.com/ec/app/apoyai/id6786046475

37. Digital Helth Technology Limitada. ApoyAI [mobile application listing]. Google Play [Internet]. Available from: https://play.google.com/store/apps/details?id=com.apoyai.app

38. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Front Public Health. 2017;5:258. doi:10.3389/fpubh.2017.00258.

39. Jang A, Hwang SK, Padhye NS, Meininger JC. Effects of cognitive behavior therapy on heart rate variability in young females with constipation-predominant irritable bowel syndrome: a parallel-group trial. J Neurogastroenterol Motil. 2017;23(3):435-445. doi:10.5056/jnm17017.

40. Johannesson E, Simrén M, Strid H, Bajor A, Sadik R. Physical activity improves symptoms in irritable bowel syndrome: a randomized controlled trial. Am J Gastroenterol. 2011;106(5):915-922. doi:10.1038/ajg.2010.480.