ORIGINAL RESEARCH

# Design and validation of a patient-reported outcome measure scale in acute pancreatitis: the PAN-PROMISE study

Enrique de-Madaria ,[1] Claudia Sánchez-Marin,[1,2] Irene Carrillo,[3] Santhi Swaroop Vege,[4] Serge Chooklin ,[5] Andriy Bilyak,[5] Rafael Mejuto,[6] Violeta Mauriz,[6] Peter Hegyi ,[7,8] Katalin Márta,[7,8] Ayesha Kamal ,[9] Eugenia Lauret-Braña,[10] Sorin T Barbu,[11] Vitor Nunes,[12] M Lourdes Ruiz-Rebollo,[13] Guillermo García-Rayado,[14] Edgard E Lozada-Hernandez,[15] Jorge Pereira,[16] Ionut Negoi,[17] Silvia Espina,[18] Marcus Hollenbach ,[19] Andrey Litvin,[20] Federico Bolado-Concejo,[21] Rómulo D Vargas,[22] Isabel Pascual-Moreno,[23] Vikesh K Singh,[9] José J Mira[3,24]

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2020-320729).

For numbered affiliations see end of article.

**Correspondence to**
Dr Enrique de-Madaria, Pancreatic Unit, Gastroenterology Department, Alicante University General Hospital, Alicante Institute for Health and Biomedical Research (ISABIAL), Alicante 03010, Spain; madaria@hotmail.com

Ed-M and CS-M are joint first authors.

1) preliminary results: European Pancreatic Club Meeting, Bergen, Norway (June 2019); 2) final results: United European Gastroenterology week, Barcelona, Spain (October 2019)

Received 21 January 2020
Revised 14 March 2020
Accepted 17 March 2020



**To cite:** de-Madaria E, Sánchez-Marin C, Carrillo I, *et al*. *Gut* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/gutjnl-2020-320729

## ABSTRACT

**Objective** This study aimed to develop and validate a patient-reported outcome measure (PROM) in acute pancreatitis (AP) as an endpoint centred on the patient.

**Design** A PROM instrument (PAtieNt-rePoRted OutcoMe scale in acute pancreatItis, an international proSpEctive cohort study, PAN-PROMISE scale) was designed based on the opinion of patients, professionals and an expert panel. The scale was validated in an international multicentre prospective cohort study, describing the severity of AP and quality of life at 15 days after discharge as the main variables for validation. The COSMIN (COnsensus-based Standards for the selection of health status Measurement INstruments) methodology was applied. Both the design and validation stages considered the content and face validity of this new instrument; the metric properties of the different items, reliability (reproducibility and internal consistence), the construct, structural and criterion validity, responsiveness and interpretability of this scale.

**Results** PAN-PROMISE consists of a seven-item scale based on the symptoms that cause the most discomfort and concern to patients with AP. The validation cohort involved 15 countries, 524 patients. The intensity of symptoms changed from higher values during the first 24 hours to lower values at discharge and 15 days thereafter. Items converged into a unidimensional ordinal scale with good fit indices. Internal consistency and split-half reliability at discharge were adequate. Reproducibility was confirmed using test–retest reliability and comparing the PAN-PROMISE score at discharge and 15 days after discharge. Evidence is also provided for the convergent-discriminant and empirical validity of the scale.

**Conclusion** The PAN-PROMISE scale is a useful tool to be used as an endpoint in clinical trials, and to quantify patient well-being during the hospital admission and follow-up.

**Trial registration number** NCT03650062

## Significance of this study

**What is already known on this subject?**

► Current endpoints in acute pancreatitis are suboptimal, as clinically relevant outcomes are infrequent, and the opinion of the patients, the centre of the healthcare effort, has not been considered before.

**What are the new findings?**

► Based on the opinion of patients, professionals and a panel of experts, a seven-symptom scale was developed and validated in an international multicentre prospective cohort study.

**How might it impact on clinical practice in the foreseeable future?**

► This is the first patient-reported outcome measurement scale in acute pancreatitis, a new tool to be used as a primary or secondary endpoint in clinical trials and to quantify patient well-being in clinical practice.

## INTRODUCTION

Acute pancreatitis (AP) is a frequent cause of hospital admission.[1] Most cases of AP have an uneventful course (mild AP), but the approximately one-third have local and/or systemic complications which are clearly associated with increased morbidity and mortality.[2–4] In the last decades, there has been an unsuccessful international effort to look for new treatments to improve the natural course of AP, so it is important to encourage the development of clinical trials to find an effective treatment for this disease. Unfortunately, there is a problem regarding this endeavour: the absence of appropriate outcome variables. The recent 'Initial Medical Treatment of AP: AGA Institute Technical Review'[5] considers death, single or multiple persistent organ failure

(>48 hours), and infected pancreatic and/or peripancreatic necrosis as clinical outcomes of importance in AP.[6] Data from a recent prospective nationwide multicentre study showed that the proportion of patients having those outcomes are 4%, 7% and 4%, respectively,[2] so clinical trials aiming to detect a reduction in such rare events would need the recruitment of thousands of patients, which may not be feasible. For this reason, new validated outcomes are needed.[7] On the other hand, the opinion of the patients, who are the centre of the healthcare effort, must be considered. There is a low level of agreement between the impact of disease on functional status from the patients' and physician's perspective.[8] Healthcare systems aiming to achieve a person-centred coordinated care should systematically measure patient satisfaction with health service (patient-reported experience measure) and the outcome associated with it (patient-reported outcome measure, PROM).[9] PROMs are validated instruments that patients complete reporting their status of health condition, without interpretation of the patient's response by a clinician.[10–13] These instruments are being reported with increasing frequency in the recent years for their ability to bridge the gap between the perceptions of the clinician and patients.[14] PROMs are used to monitor individual patient outcomes, gathering information directly from patients about their symptoms.[13] This information is then used to adjust treatment and care and to achieve better results, enhance adherence, increase patient satisfaction and rethink how healthcare is organised and delivered.[12 14–16] PROMs have been usually designed for patients with chronic conditions, or for patients undergoing surgical procedures, and are rarely reported in acute diseases. Finally, PROMs are gaining importance as a tool to design outcome variables for the clinical trials.[11] Studies designed to assess the efficacy for new treatments for AP should include PROMS as an important outcome.

This study aimed to design a PROM in AP (the PAtieNt-rePoRted OutcoMe scale in acute pancreatItis, an international proSpEctive cohort study, PAN-PROMISE Scale) and to validate it in an international prospective cohort of patients.

## METHODS

PAN-PROMISE was designed and validated following the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) methodology (COSMIN Study Design Checklist for patient-reported outcome measurement instruments is available in online supplementary material 1).[17 18]

The development of PAN-PROMISE was conducted in two phases (design and analysis of its metric properties). They included: (1) operational definition of the feature to be measured, validity of the design considering content, face and cross-cultural validity of this new instrument; (2) the analysis of the metric properties of the different items, reliability (internal consistence and reproducibility), the construct, structural and criterion validity, and the analysis of the responsiveness and interpretability of this scale.

The project was carried out in accordance with the standards of good clinical practice and the international ethical principles applicable to medical research in humans (Declaration of Helsinki).[19] Written informed consent was required for patients to participate in the study in both the design and validation phase.

### Phase I: design of the PAN-PROMISE scale

#### Operational definition

PAN-PROMISE was designed as a standardised, validated instrument completed by patients with AP to measure their perception of their functional well-being and health status. This new PROM was designed involving both patients and professionals, as their views are complementary.[20]

#### Content validity

A qualitative technique was applied to assure content validity. Three nominal groups were conducted, two of them included only patients recovered from an episode of AP: (A) seven patients from Alicante University General Hospital, and (B) seven patients from Valencia's Clinic Hospital, Valencia, Spain. AP was defined according to the revision of the Atlanta Classification (RAC)[3] (see detailed description in the section 'Variables' in phase II methodology). The third group included professionals (six gastroenterologists, one internist and one nurse), from the Alicante region (Spain) with more than 5 years experience in treating patients with AP. Patients were enrolled considering voluntariness, gender and severity[3] of their symptoms, and overall included seven women and seven men, four mild, four moderate and six severe AP. Participant professionals were recruited considering voluntariness and their clinical experience.

Patients were asked about the symptoms that caused them the most discomfort and concern at four specific time points: before receiving treatment for AP, during hospital admission, discharge and after discharge. Similarly, professionals were asked about the symptoms that, according to their experience, cause the most discomfort and concern to their patients at the same time points. Nominal groups were conducted by psychologists experienced in qualitative research (IC and JJM), first by collecting individual views and, second, promoting debate about these initial views. Debate continued until no new information emerged and saturation of information was reached. An analysis of the consistency (intragroup and among groups, triangulation) of those ideas was also carried out. The perspectives of patients and professionals were compared.

#### Face validity

This first list of symptoms was shared with an international group of nine experts (gastroenterologists and surgeons) in pancreatology (see the Acknowledgements) experienced in clinical care and research in the field of AP, through an online application to determine to what extent these experiences were shared in frequency and intensity by patients in their different countries. An important task of these international experts was to detect, comment and correct possible cultural differences. Based on consensus among those international experts, the PAN-PROMISE instrument was generated.

The understanding of items was evaluated by three patients who had been recently discharged from Alicante University General Hospital after an episode of AP. Their opinions regarding the relevance of those symptoms were also considered.

#### Cross-cultural validity

The PAN-PROMISE scale was initially developed in Spanish and then it was translated into English following the forward-back translation method as recommended by WHO.[21] The wording of the items considered the national language particularities of countries involved in this study. We applied the same procedure to translate the scale to other languages.

### Phase II: validation of the PAN-PROMISE scale

For the validation of the scale, an international multicentre prospective cohort study was performed.

### Patients

Patients with AP, between 18 and 80 years of age, with a Karnofsky performance status[22] previous to the episode of AP equal or higher than 80 were included after informed consent. For the diagnosis of AP, the presence of at least two of the following criteria were required: (A) typical upper abdominal pain, (B) increase in serum amylase and/or lipase above three times the upper limit of normal and (C) imaging compatible with AP.[3]

Exclusion criteria were as follows: more than one previous episode of AP; chronic pancreatitis; time between onset of symptoms and presentation in the emergency room greater than 48 hours; recruitment more than 24 hours after presentation in the emergency room; inability to understand the instructions of the study or communicate with the researchers (severe congenital or acquired intellectual deficit); presence of diseases or conditions different from AP that may interfere with the scale, for example, other causes of abdominal pain (especially acute cholecystitis), obstruction of the digestive tract (peptic pyloric stenosis, gastrointestinal anastomotic stenosis, diabetic gastroparesis, gastrointestinal neoplasia…), nausea–vomiting (brain tumour, chemotherapy…) or weakness (pre-existing anaemia with haemoglobin <9 g/dL, heart failure or respiratory insufficiency associated with minimal effort dyspnoea, or domiciliary treatment with $O_2$, advanced neoplasms or other debilitating diseases).

Criteria for exclusion of analysis: diagnosis after inclusion in the study of previous or new diseases, different from AP, with a potential impact on the PAN-PROMISE scale, for example, acute cholecystitis, severe sepsis, chronic pancreatitis, neoplasia.

A sample size of at least 384 AP patients was determined, considering the most unfavourable option in the calculation of a positive assessment proportion for a p=q=0.50, defining a 95% confidence level and accepting a maximum error of 5%, expected response rate higher than 80%. Stratified sampling using severity categories of AP defined by the RAC[3] was applied (expected to be approximately 65% mild, 28% moderate and 7% severe, according to a previous multicentre prospective cohort study.[2] Subjects who did not answer 85% of the questions were excluded.

**Table 1** Baseline characteristics and outcomes of the validation study

| Characteristics and outcomes | |
|---|---|
| n | 524 |
| Age, median (P25–P75) years | 55 (41–66) |
| Male sex, n (%) | 281 (53.6) |
| BMI, median (P25–P75) kg/m$^2$ | 28.1 (25–32.3) |
| CCI, median (P25–P75) points | 0 (0–1) |
| AP episode, n (%) | |
| First | 446 (85.1) |
| Second | 78 (14.9) |
| Aetiology, n (%) | |
| Gallstones | 286 (54.6) |
| Alcohol | 110 (21) |
| Idiopathic | 56 (10.7) |
| Other | 72 (13.7) |
| SIRS criteria, n (%) | |
| No | 360 (68.7) |
| Transient SIRS criteria (≤48 hours) | 107 (20.4) |
| Persistent SIRS criteria (>48 hours) | 57 (10.9) |
| C reactive protein serum levels, median (P25–P75) mg/L | |
| At day 2 | 100 (35–191) |
| At day 3 | 105 (40–218) |
| Organ failure, n (%) | |
| No OF | 448 (85.5) |
| Transient OF (≤48 hours) | 30 (5.7) |
| Persistent OF (>48 hours) | 46 (8.8) |
| Local complications, n (%) | |
| No local complications | 341 (65.1) |
| APFC | 74 (14.1) |
| Peri(pancreatic) necrosis | 109 (20.8) |
| ICU admission, n (%) | 53 (10.1) |
| Nutritional support, n (%) | 141 (26.9) |
| Invasive treatment, n (%) | 41 (7.8) |
| Hospital stay, median (P25–P75) days | 7 (5–11) |
| Readmission from discharge to day 15±2 after discharge, n (%) | 18 (3.4) |
| Cholecystectomy during index admission, n (%) | 73 (13.9) |
| Mortality, n (%) | 15 (2.9) |
| Severity, n (%) | |
| Mild | 325 (62) |
| Moderately severe | 153 (29.2) |
| Severe | 46 (8.8) |

P25–P75: 25 and 75 percentiles.
SIRS criteria: 2 or more systemic inflammatory response syndrome criteria.
Peri (pancreatic) necrosis: pancreatic and/or peripancreatic necrosis.
AP, acute pancreatitis; APFC, acute peripancreatic fluid collections; BMI, body mass index; CCI, Charlson Comorbidity Index; ICU, intensive care unit; OF, organ failure.

### Variables

All local and systemic complications of AP were defined according to the RAC,[3] including pancreatic necrosis, peripancreatic necrosis, acute peripancreatic fluid collections, exacerbation of previous comorbidity, organ failure and subtypes of organ failure (≤48 hours duration=transient and >48 hours duration=persistent organ failure). Invasive treatment was defined as any of the following: percutaneous or endoscopic drainage, surgical or endoscopic necrosectomy or endoscopic retrograde cholangiopancreatography due to other causes than choledocholithiasis (eg, main pancreatic duct disruption). Severity was based on the RAC: severe AP is defined by the presence of persistent organ failure regardless of other complications, moderately severe AP by the presence of local complications, exacerbation of previous comorbidity and/or transient organ failure, finally mild AP is defined by the absence of complications and organ failure.[3] Quality of life was assessed by means of the EORTC QLQ-C30 questionnaire.[23]

### Data acquisition

Local collaborators were in charge of recruiting patients and obtaining the data required for the study. Patients were recruited on the first 24 hours after presentation in the emergency room. The PAN-PROMISE scale was measured in the first 24 hours after presentation, at day 2 after presentation, day 5, day 7, at discharge, and 15 (±2) days after discharge (by phone call or during outpatient clinic visit). EORTC QLQ-C30 was obtained once, at 15 (±2) days after discharge.

Study data were collected and managed using the online electronic case report form tool Research Electronic Data Capture (REDCap), hosted at Asociación Española de Gastroenterología (AEG; www.aegastro.es). AEG is a non-profit scientific and medical association focused on gastroenterology. This service was provided free of charge, with the sole aim of promoting independent investigator-driven research. REDCap is a secure,

**Table 2** Cronbach's alpha, Mc Donald's omega and split-half reliability tests

| | | First 24 hours | Day 2 | Discharge | 15±2 days after discharge |
|---|---|---|---|---|---|
| Overall (n=524) | Cronbach's alpha | 0.73 | 0.82 | 0.73 | 0.77 |
| | Mc Donald's omega | 0.81 | 0.88 | 0.79 | 0.87 |
| | Split-half correlation between forms KR20 (odd vs even) | 0.56 | 0.67 | 0.59 | 0.65 |
| | Guttman Split-Half Coefficient | 0.71 | 0.79 | 0.74 | 0.79 |
| Mild AP (n=323) | Cronbach's alpha | 0.72 | 0.81 | 0.73 | 0.77 |
| | Mc Donald's omega | 0.80 | 0.87 | 0.78 | 0.84 |
| | Split-half correlation between forms KR20 (odd vs even) | 0.54 | 0.64 | 0.59 | 0.67 |
| | Guttman split-half coefficient | 0.69 | 0.77 | 0.74 | 0.80 |
| Moderately severe +severe (n=201) | Cronbach's alpha | 0.69 | 0.75 | 0.72 | 0.76 |
| | Mc Donald's omega | 0.75 | 0.82 | 0.81 | 0.87 |
| | Split-half correlation between forms KR20 (odd vs even) | 0.52 | 0.61 | 0.62 | 0.62 |
| | Guttman split-half coefficient | 0.68 | 0.75 | 0.75 | 0.76 |

AP, acute pancreatitis; KR20, Kuder-Richardson Formula 20.

web-based application designed to support data capture for research studies.[24] Data were anonymised in REDCap.

## Data analysis

### Metric proprieties of the items

The floor and ceiling effects of each item were analysed individually. The item-total correlation was also analysed to characterise the metric proprieties of the elements, excluding those items with low correlations. A minimum of 0.35 Pearson's correlation was considered as acceptable.

### Reliability

Cronbach's alpha and McDonald's omega were used to estimate internal consistency. Split-half reliability was also estimated using the Spearman-Brown and the Kuder-Richardson coefficients. Items were randomly sorted and they were split into two parts before applying this statistic. A value greater than 0.70 was considered acceptable for both statistics. Test–retest reliability (reproducibility) was also applied comparing the PAN-PROMISE score at discharge and 15±2 days after discharge using t-test. To test possible cultural differences that may affect the consistence of the PAN-PROMISE scale, we compared patients from Western Europe (Spain, Greece, Italy, Portugal and Germany) with those of Eastern Europe (Ukraine, Romania, Hungary, Russia, Bulgaria and Poland).

### Construct validity

Exploratory factor analysis to determine the factorial structure of this instrument was conducted using the principal components technique, followed by Varimax rotation. Eigen values greater than 0.40 and factor loading greater than 0.5 were considered an acceptable level of missing data. The suitability for this factorial analysis of the interitem correlation matrix was calculated using the Bartlett's sphericity test and the Kaiser-Meyer-Olkin (KMO) test at 24 and 48 hours from hospitalisation and discharge.

To further investigate the construct validity, a unidimensional scale was hypothesised and tested applying the parallel analysis based on the minimum rank factor analysis of 500 random correlation matrices obtained by the permutation of the raw data. Factor application was used to run this analysis. Additionally, a confirmatory factor analysis (CFA) was used to explore the underlying structure and to attempt to reduce the overall number of items into latent factors based on commonalities within the data. Several fit indices were selected in order to test which CFA model best represented the dataset: Comparative Fit Index (CFI), Goodness of Fit Index (GFI), adjusted GFI (AGFI), standardised root mean-square residual (RMSEA). Values greater than 0.90 for CFI and AGFI, greater than 0.9% for GFI and less than 0.8 for RMSEA were considered to be indicators of good fitting model. To test the models, a lower $X^2$ value indicates a better fit, given an equal number of degrees of freedom (df).

### Structural validity

The convergent-discriminant method was used considering that increasing scores on the PAN-PROMISE scale were expected to be associated to decreased quality of life measured by the EORTC QLQ-C30 (V.3). Analysis of variance was used to test whether the different severity categories (the main outcome to validate the scale) had different scores on the PAN-PROMISE scale: severe greater than moderate and mild, moderate greater than mild. Accuracy to predict moderate to severe disease was investigated by means of receiver operating characteristic (ROC) curve, sensitivity and specificity.

**Table 3** Test–retest reliability (reproducibility analysis, results at discharge and 15±2 days after discharge)

| | Mean of the difference | SD | 95% CI | | T-test | P value |
|---|---|---|---|---|---|---|
| Pain, especially in the abdomen, chest or back | 0.10 | 1.62 | −0.04 | 0.24 | 1.42 | 0.155 |
| Abdominal distention (bloating, sensation of excess gas) | 0.18 | 1.81 | 0.02 | 0.34 | 2.26 | 0.024 |
| Difficulty eating, sensation of food being stuck in the stomach | 0.06 | 1.46 | −0.06 | 0.19 | 0.97 | 0.331 |
| Difficulty with bowel movements (constipation or straining on bowel movements) | 0.12 | 2.10 | −0.06 | 0.31 | 1.33 | 0.184 |
| Nausea and/or vomiting | 0.01 | 0.99 | −0.08 | 0.09 | 0.18 | 0.857 |
| Thirst | 0.19 | 1.79 | 0.04 | 0.35 | 2.41 | 0.016 |
| Weakness, lack of energy, fatigue, difficulty moving | 0.15 | 2.09 | −0.03 | 0.33 | 1.59 | 0.112 |

CI, Confidence Interval; SD, Standard Deviation.

**Table 4** Fit indexes for the PAN-PROMISE scale

| Fit index (reference value considered as acceptable) | Two-factor model | One-factor model |
|---|---|---|
| Comparative Fit Index | 0.90 | 0.93 |
| Jöreskog-Sörbom's Goodness of Fit Index (GFI) | 0.96 | 0.96 |
| Adjusted GFI | 0.92 | 0.92 |
| Standardised root mean-square residual | 0.05 | 0.04 |

PAN-PROMISE, PAtieNt-rePoRted OutcoMe scale in acute pancreatItis, an International proSpEctive cohort study.

Linear regression was also used to estimate the convergent-discriminant validity of the PAN-PROMISE scale at the first 24 hours from admission, at day 2, day 5 and at discharge.

### Criterion validity

Lineal regression was used to determine the prediction of hospital stay by the PAN-PROMISE scale. The relationship between PAN-PROMISE score and severity (mild vs moderate to severe AP) was analysed by means of binary logistic regression.

### Responsiveness

The scores of the PAN-PROMISE scale from the first 24 hours to discharge were compared to determine the ability of this scale to detect change over time in the construct to be measured.

### Interpretability

A set of clinicians assessed the meaning of the PAN-PROMISE score as an outcome in the course of the health care received by AP patients

## RESULTS

### Phase I: design of the PAN-PROMISE scale

#### Face and content validity

Patients and professionals coincided in listing relevant symptoms in AP, but the perspective of professionals regarding the intensity-order of importance of these symptoms differed from patients, see online supplementary material 2.

With the information gathered in the nominal groups, a 20-item first version of the PAN-PROMISE scale (online supplementary material 3) was elaborated. This list was shared with the nine international experts in pancreatology, and based on consensus among them, a set of seven reactive items for the PAN-PROMISE instrument was generated:

Each item is scored from 0 to 10 (worst score in the last 24 hours, 0: none, 10: the highest possible intensity)

1. Pain, especially in the abdomen, chest or back.
2. Abdominal distention (bloating, sensation of excess gas).
3. Difficulty eating, sensation of food being stuck in the stomach.
4. Difficulty with bowel movements (constipation or straining on bowel movements).
5. Nausea and/or vomiting.
6. Thirst.
7. Weakness, lack of energy, fatigue, difficulty moving.

PAN-PROMISE total score was the sum of the score assigned to each of the seven items.

#### Cross-cultural validity

The PAN-PROMISE scale was translated to Bulgarian, Chinese, English, French, German, Greek, Hungarian, Italian, Korean, Polish, Portuguese, Romanian, Russian, Turkish, Ukrainian and Urdu. Online supplementary material 4 includes all available versions of this instrument.

### Phase II: metric properties of the PAN-PROMISE scale

From May 2017 to November 2018, 524 patients from 29 centres (15 countries) were recruited (see online supplementary material 5 for information regarding languages, countries and centres). The PAN-PROMISE score was not available (missing data) for 0 patients at the first 24 hours, 55 (10.5%) at day 2, 28 (5.3%) at day 5, 151 (28.8%) at day 7, 16 (3.1%) at discharge and 17 (3.2%) at 15±2 days from discharge. EORTC QLQ-C30 scale was missing in 18 (3.4%) patients at 15±2 days from discharge. Baseline characteristics and outcomes are shown in table 1.

Outcomes according to severity of AP are displayed in online supplementary material 6. As expected, increasing severity was associated with worse outcomes.

#### Consistency and reliability

A floor or ceiling effect was not identified in any item. The item-total correlations during the first 24 hours varied from 0.41 to 0.49, at discharge between 0.36 and 0.56 (except 'nausea and/or vomiting' which had a correlation of 0.26) and at 15±2 days in the range of 0.44 and 0.61.

The results of the PAN-PROMISE scale at different time points in Cronbach's alpha test, McDonald's omega and split-half method (overall and stratified by severity) are displayed in table 2. Those results at discharge and 15±2 days after discharge remained similar (reproducibility) with the exception of thirst and abdominal distension with improved from discharge to 15±2 days after discharge (table 3). Consistency was found in both Eastern and Western European countries (online supplementary material 7).

#### Construct validity

The interitem correlation matrix was suitable for factorial analysis at 24 hours from hospitalisation, based on the Bartlett's sphericity test ($\chi^2$=625.0; df=21; p=0.0001), and the KMO test (KMO=0.772); at 48 hours from hospitalisation, the Bartlett's Sphericity test ($\chi^2$=1036.6; df=21; p=0.0001), and the KMO test (KMO=0.834); and at discharge, the Bartlett's sphericity test ($\chi^2$=632.1; df=21; p=0.0001) and the KMO test (KMO=0.784).

Regarding construct validity, the parallel analysis recommended a one-factor dimension as a unidimensional scale achieved the best fit indexes, see table 4, and was confirmed by stratifying by severity, online supplementary material 8.

#### Structural validity

Correlations between the Quality of Life EORTC QLQ-C30 scale and PAN-PROMISE scale at 15±2 days after discharge are shown in table 5. It confirmed that the PAN-PROMISE scale had adequate discriminant and convergent validity.

#### Criterion validity

PAN-PROMISE score at day 2 was associated to:

A. Hospital stay (beta 0.33, p=0.0001, 95% CI 0.17 to 0.36); a higher score was linked to a longer hospital stay.
B. Severity (Wald=31.47, beta 1.06, 95% CI 1.04 to 1.08); a higher score was associated to moderate to severe disease.

**Table 5** Convergent-discriminant analysis

| | Total score | Pain | Abdominal distention | Difficulty eating | Difficulty with bowel movements | Nausea and/or vomiting | Thirst | Weakness |
|---|---|---|---|---|---|---|---|---|
| **Eortc functional scales** | −0.48** | −0.33** | −0.30** | −0.28** | −0.28** | −0.25** | −0.27** | −0.45** |
| Global health status | −0.33** | −0.28** | −0.23** | −0.21** | −0.17** | −0.18** | −0.15** | −0.30** |
| Physical functioning | −0.45** | −0.30** | −0.30** | −0.27** | −0.26** | −0.22** | −0.25** | −0.43** |
| Role functioning | −0.35** | −0.33** | −0.25** | −0.23** | −0.15** | −0.16** | −0.10* | −0.37** |
| Emotional functioning | −0.43** | −0.26** | −0.27** | −0.23** | −0.31** | −0.24** | −0.28** | −0.34** |
| Cognitive functioning | −0.23** | −0.11* | −0.13** | −0.07 | −0.08 | −0.06 | −0.23** | −0.26** |
| Social functioning | −0.20** | −0.14** | −0.08 | −0.13** | −0.11* | −0.18** | −0.13** | −0.17** |
| Financial difficulties | 0.20** | 0.12** | 0.14** | 0.14** | 0.22** | 0.19** | 0.13** | 0.05 |
| **Eortc symptom scales** | 0.55** | 0.36** | 0.34** | 0.39** | 0.39** | 0.28** | 0.30** | 0.47** |
| Fatigue | 0.56** | 0.32** | 0.35** | 0.35** | 0.29** | 0.19** | 0.28** | 0.64** |
| Nausea vomiting | 0.26** | 0.15** | 0.15** | 0.21** | 0.15** | 0.41** | 0.17** | 0.12** |
| Pain | 0.40** | 0.51** | 0.24** | 0.26** | 0.18** | 0.22** | 0.19** | 0.30** |
| Dyspnoea | 0.19** | 0.13** | 0.12** | 0.13** | 0.07 | 0.13** | 0.08 | 0.20** |
| Insomnia | 0.36** | 0.20** | 0.21** | 0.26** | 0.37** | 0.12** | 0.27** | 0.21** |
| Appetite loss | 0.37** | 0.12** | 0.22** | 0.38** | 0.22** | 0.22** | 0.17** | 0.36** |
| Constipation | 0.34** | 0.05 | 0.23** | 0.26** | 0.62** | 0.06 | 0.16** | 0.12** |
| Diarrhoea | 0.01 | 0.08 | 0.01 | −0.02 | −0.03 | 0.01 | −0.03 | 0.04 |

Correlations between EORTC QLQ-C30 scale and PAN-PROMISE scale (15±2 days after discharge).
*P<0.05; **P<0.01.
PAN-PROMISE, PAtieNt-rePoRted OutcoMe scale in acute pancreatItis, an intErnational proSpEctive cohort study.

## Responsiveness

The intensity of symptoms changed from higher values during the first 24 hours to lower values at discharge and subsequently, at 15±2 days from discharge, see table 6 and online supplementary material 9.

PAN-PROMISE score at day 2, discharge and 15±2 days after discharge stratified by outcomes of AP is shown in table 7. A significantly higher PAN-PROMISE score was reported for patients who needed invasive treatment (day 2, discharge and 15±2 days after discharge), nutritional support (day 2 and discharge), intensive care unit (day 2), organ failure (day 2) and had systemic inflammatory response syndrome (SIRS) (day 2). Furthermore, mean PAN-PROMISE score at day 2 was higher in persistent OF than transient OF, and transient OF than no OF. Similarly, mean score at day 2 was higher in persistent SIRS than transient SIRS, and transient SIRS higher than no SIRS (table 7). The areas under the ROC curve (AUC) for detecting moderate to severe disease, best cut-off points and their sensitivity and

**Table 6** PAN-PROMISE mean scores at different time points from the first 24 hours to 15±2 days after discharge

| | | 24 hours | Day 2 | Day 5 | Day 7 | Discharge | 15±2 days after discharge |
|---|---|---|---|---|---|---|---|
| Pain | Mild | 7.33 | 3.10 | 1.37 | 0.92 | 0.75 | 0.59 |
| | Moderate to severe | 7.95 | 5.11 | 2.64 | 1.63 | 0.72 | 0.70 |
| | Total | 7.57 | 3.87 | 1.85 | 1.23 | 0.74 | 0.63 |
| Abdominal distension | Mild | 4.32 | 2.18 | 1.30 | 0.95 | 0.91 | 0.60 |
| | Moderate to severe | 5.55 | 4.15 | 2.48 | 1.73 | 0.79 | 0.79 |
| | Total | 4.79 | 2.94 | 1.74 | 1.30 | 0.86 | 0.67 |
| Difficulty eating | Mild | 3.50 | 1.56 | 0.91 | 0.49 | 0.47 | 0.34 |
| | Moderate to severe | 5.48 | 3.53 | 2.20 | 1.27 | 0.43 | 0.53 |
| | Total | 4.26 | 2.32 | 1.39 | 0.84 | 0.46 | 0.41 |
| Constipation | Mild | 2.00 | 1.60 | 1.07 | 0.76 | 0.81 | 0.59 |
| | Moderate to severe | 3.69 | 3.04 | 1.89 | 1.04 | 0.70 | 0.69 |
| | Total | 2.65 | 2.15 | 1.37 | 0.88 | 0.77 | 0.63 |
| Nausea and/or vomiting | Mild | 4.25 | 1.22 | 0.42 | 0.22 | 0.20 | 0.15 |
| | Moderate to severe | 5.20 | 2.25 | 0.82 | 0.51 | 0.09 | 0.19 |
| | Total | 4.62 | 1.62 | 0.57 | 0.35 | 0.16 | 0.16 |
| Thirst | Mild | 4.04 | 2.71 | 1.38 | 0.95 | 0.78 | 0.54 |
| | Moderate to severe | 5.12 | 3.97 | 2.09 | 1.19 | 0.62 | 0.62 |
| | Total | 4.45 | 3.20 | 1.64 | 1.05 | 0.73 | 0.57 |
| Weakness | Mild | 4.17 | 2.66 | 2.08 | 1.79 | 1.45 | 1.27 |
| | Moderate to severe | 5.25 | 4.36 | 3.01 | 2.20 | 1.58 | 1.47 |
| | Total | 4.58 | 3.31 | 2.43 | 1.97 | 1.49 | 1.34 |
| Total | Mild | 29.61 | 15.03 | 8.25 | 6.07 | 5.36 | 4.09 |
| | Moderate to severe | 38.24 | 26.42 | 15.13 | 9.58 | 4.95 | 5 |
| | Total | 32.91 | 19.42 | 11 | 7.61 | 5.21 | 4.43 |

PAN-PROMISE, PAtieNt-rePoRted OutcoMe scale in acute pancreatItis, an intErnational proSpEctive cohort study.

**Table 7** Convergent-discriminant analysis

| | | PAN-PROMISE score at day 2 | | | PAN-PROMISE score at discharge | | | PAN-PROMISE score at 15±2 days after discharge | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | n | Mean (SD) | P value | n | Mean (SD) | P value | n | Mean (SD) | P value |
| Local complications*† | Yes | 169 | 26.7 (13.7) | 0.732 | 169 | 4.9 (6.2) | 0.196 | 169 | 5.0 (6.7) | 0.911 |
| | No | 12 | 25.3 (12.6) | | 17 | 7.0 (6.3) | | 17 | 5.2 (5.6) | |
| Invasive treatment*‡ | Yes | 33 | 35.0 (12.4) | 0.000 | 33 | 6.9 (7.4) | 0.038 | 33 | 6.8 (7.8) | 0.034 |
| | No | 148 | 24.8 (13.2) | | 153 | 4.8 (5.9) | | 153 | 4.7 (6.3) | |
| Nutritional support† | Yes | 137 | 23.7 (15.9) | 0.000 | 129 | 4.2 (5.2) | 0.016 | 128 | 4.2 (5.1) | 0.682 |
| | No | 332 | 17.9 (13.5) | | 379 | 5.6 (7.3) | | 379 | 4.5 (7.1) | |
| Intensive care unit admission*‡ | Yes | 48 | 34.6 (12.6) | 0.000 | 37 | 5.8 (6.7) | 0.266 | 37 | 5.6 (5.8) | 0.129 |
| | No | 133 | 23.8 (12.8) | | 149 | 5.0 (6.1) | | 149 | 4.9 (6.8) | |
| Organ failure*§ | No | 116 | 23.6 (12.9) | 0.000 | 125 | 4.5 (5.9) | 0.167 | 125 | 5.1 (7.1) | 0.972 |
| | Transient (≤48 hours) | 26 | 26.3 (13.7) | | 30 | 6.7 (6.9) | | 30 | 4.8 (5.4) | |
| | Persistent (>48 hours) | 39 | 36.0 (11.4) | | 31 | 6.0 (6.8) | | 31 | 5.1 (5.7) | |
| Systemic inflammatory response syndrome*§ | No | 84 | 22.8 (12.8) | 0.000 | 89 | 4.5 (5.3) | 0.272 | 89 | 4.7 (7.4) | 0.694 |
| | ≤48 hours | 48 | 25.8 (12.3) | | 51 | 6.2 (7.5) | | 51 | 5.0 (5.4) | |
| | >48 hours | 47 | 34.5 (13.5) | | 45 | 5.3 (6.4) | | 45 | 5.8 (6.3) | |

PAN-PROMISE score at day 2, discharge and 15±2 days after discharge and outcomes of acute pancreatitis.
*Selection of moderate and severe cases.
†t-test.
‡Mann-Whitney U test.
§One-way ANOVA.
ANOVA, analysis of variance ; PAN-PROMISE, PAtieNt-rePoRted OutcoMe scale in acute pancreatItis, an intErnational proSpEctive cohort study.

specificity are shown in table 8. The best AUC and best balance of sensitivity and specificity were on day 2. At discharge and 15±2 days after discharge, symptoms were similar in both mild and moderate to severe disease, with AUC close to 0.500.

### Interpretability

The scores on the scale were considered to have a direct dual purpose. First, they provide an additional criterion for the outcome of the intervention and may guide the decision on when to discharge from hospital. Second, they provide an average of the therapeutic utility from the patient's perspective. This measure could be useful both for decisions regarding clinical care and for research into new treatments.

## DISCUSSION

Despite being one of the most frequent diseases of the gastrointestinal system requiring hospital admission, a specific treatment for AP remains elusive. For the development of controlled trials, relevant outcome variables are needed. Given the low frequency of events in researcher-defined clinically relevant endpoints like mortality, persistent organ failure or infection of pancreatic necrosis, which makes sample size unfeasible, different approaches can be taken. First, a surrogate variable can be chosen, like C reactive protein or incidence of systemic inflammatory response syndrome. Unfortunately, an intervention frequently has a statistically significant improvement on a surrogate marker but no measurable effect on clinically relevant outcomes.[25] Thus, surrogate outcomes are considered by many authors as inadequate.[5] A second possibility is to use composite variables that combine different important outcomes. A special type of composite variables are disease activity scores like the recently developed Pancreatitis Activity Scoring System.[7] Composite variables increase statistical efficiency, however, the interpretation of results is challenging and reporting of composite outcomes is frequently inadequate.[26] Remarkably, all that effort to develop and use different endpoints has been led only by researchers, without input from the patients, who are the centre of healthcare.[12] PAN-PROMISE aims to be an outcome variable with intrinsic importance, as it is focused on the most important symptoms for the patient with AP in terms of discomfort and concern, being the strengths of this project the following: (A) It was developed considering the patient's point of view together with professionals with experience in managing this disease, following the recommendations regarding this kind of studies[20] being an holistic approach to symptoms and concerns; (B) A panel with some of the most important researchers in AP were involved in its development (see the Acknowledgements) and (C) it has been validated in a researcher-driven international (15 countries) multicentre (29 centres) prospective cohort of patients specifically designed for this purpose.

The PAN-PROMISE scale showed good consistency, reliability, reproducibility, convergent-discriminant and empirical validity, so it can be used as a primary or secondary endpoint for clinical trials aiming to investigate new treatments for AP, as this instrument will verify the effect of such treatments on the patients' well-being, exploring this important dimension of healthcare. It is also a tool for daily clinical practice, to check our patients' symptoms and improvement in an easier and 'measurable' way that conventional anamnesis. Its use does not require permission from the PAN-PROMISE team. The scale is very simple, is available in several languages (see online supplementary material 4)

**Table 8** Accuracy of the PAN-PROMISE scale for predicting moderate to severe disease

| | AUC | Best cut-off point | Sensitivity, % | Specificity, % |
|---|---|---|---|---|
| Day 1 | 0.672 | 33.5 | 69.0 | 63.5 |
| Day 2 | 0.750 | 19.5 | 69.7 | 71.8 |
| Day 5 | 0.695 | 5.5 | 76.9 | 55.4 |
| Day 7 | 0.642 | 3.5 | 68.9 | 57.1 |
| Discharge | 0.531 | 0.5 | 80.8 | 31.9 |
| 15±2 days after discharge | 0.565 | 0.5 | 71.4 | 40.8 |

AUC, area under the curve; PAN-PROMISE, PAtieNt-rePoRted OutcoMe scale in acute pancreatItis, an intErnational proSpEctive cohort study.

and can be fulfilled by the patient in 1 min or less, so repeated measures are easy to perform, it is not time-consuming for the physician and can be easily incorporated to both research and clinical practice without an increase in workload.

Interestingly, the intensity of some concerns of the patients were not expected by the PAN-PROMISE professionals (see online supplementary material 2), a gap that has been described before.[8] For example, weakness was rated by the patients as the second most concerning symptom, only after abdominal pain, and thirst was in the third position, being rated by the professionals in the 24th and 7th position, respectively. For that reason, when developing scores that include symptoms, patients have to be always involved. Thirst and abdominal distension were the only two items of the PAN-PROMISE scale that had a significant improvement between discharge and 15±2 days after discharge. Thirst is a frequent symptom in critically ill patients.[27] Moderate to severe AP is associated to fluid sequestration[28] leading to intravascular volume loss which in turn activates the renin–angiotensin–aldosterone system.[27] In addition, subtle changes in plasma osmolality induce the release of vasopressin. Both systems are involved in producing thirst. Some drugs commonly used in AP are also associated this symptom: diuretics, opioids, non-steroidal anti-inflammatory drugs and proton pump inhibitors.[27] It seems that at discharge, these complex mechanisms are still active. Abdominal distension is an unspecific symptom that presumably is also caused by a constellation of factors like collections, paralytic ileus, delayed gastric emptying, digestive tract dysfunction, bacterial overgrowth, pancreatic exocrine insufficiency… We hypothesise that some of these causes, like pancreatic exocrine insufficiency or bacterial overgrowth can be present at discharge, in fact pancreatic exocrine insufficiency may be a long-lasting sequel.[29]

According to our data, symptoms at discharge and 15±2 days after discharge are quite similar in mild versus moderate to severe disease. At the beginning of the disease, the patients have the maximum intensity of symptoms and local complications are not present; then in mild disease the patient improves quicker than in moderate to severe AP. The scale had an AUC for detecting moderate to severe disease at 24 hours of 0.672 and increased at 48 hours to 0.750. The scale yielded and AUC in other time points before discharge of 0.642–0.695. For that reason, we think that PAN-PROMISE scale at 48 hours may be a good endpoint in clinical trials addressing the early treatment of AP. Mean PAN-PROMISE score in patients with mild disease was approximately 10 points lower than patients with moderate to severe disease during the first days of hospital admission (table 6), suggesting that changes of 10 points may be a possible endpoint for future trials. Patients are discharged when symptoms have subsided, so AUC at discharge and 15±2 days after discharge are close to 0.500. It is difficult that a symptom scale can be more accurate to detect severity, as the different severity categories are defined by imaging and organ function status. It is important to emphasise that this scale is not intended to be used to predict severity by means of a cut-off point which would be different for each different time point, but to report the intensity of symptoms in a given moment, so we can compare the effect of different treatments on those symptoms.

This is a special PROM for two reasons, first, it is the first one to be designed specifically for AP. Second, it is a PROM designed for an acute disease; most PROMs have been designed for chronic conditions or to compare patient symptoms before and after an invasive procedure, so the PAN-PROMISE scale will be helpful for the future development of new PROMS for other acute diseases.

Our study had some weaknesses. The scale was derived from Spanish patients and healthcare professionals so it may be exposed to cultural peculiarities; to control this problem we had a panel of international pancreatologists to detect those possible cultural issues (including leaders in clinical research in pancreatitis from India, Turkey, Germany and USA); furthermore, we demonstrated good internal consistency in both Western and Eastern European countries (online supplementary material 7). Qualitative techniques involved a total of 14 patients. They suffered clinical situations that represented usual patient profiles in hospitals. Although the focus of the group discussion was on AP-derived symptoms, it is difficult to differentiate possible influences on patient experience associated with health beliefs or social factors. The number of patients recruited and surveyed by different centres was wide, and some centres included few patients, which may be associated with biases. Possible biases regarding the inclusion of non-consecutive and few patients are possibly corrected by the high overall number of patients recruited, more than 500. In fact, outcomes in the PAN-PROMISE cohort of patients are very similar to outcomes in a recent nationwide prospective cohort study from our group which included more than 1600 patients[2] (online supplementary material 10), so external validity is guaranteed.

## CONCLUSIONS

The PAN-PROMISE scale is a useful tool to be used as an endpoint in clinical trials, and to quantify patient well-being during the hospital admission and follow-up.

**Author affiliations**
[1]Pancreatic Unit, Gastroenterology Department, Alicante University General Hospital, Alicante Institute for Health and Biomedical Research (ISABIAL), Alicante, Spain
[2]Department of Gastroenterology and Hepatology, Marqués de Valdecilla University Hospital, Santander, Spain
[3]Department of Health Psychology, Universidad Miguel Hernández, Elche, Spain
[4]Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, Minnesota, USA
[5]Department of Surgery, Lviv Regional Clinical Hospital, Lviv, Ukraine
[6]Department of Gastroenterology and Hepatology, University Hospital of Santiago de Compostela, Santiago de Compostela, Spain
[7]Institute for Translational Medicine, Szentágothai Research Centre, University of Pécs Medical School, Pécs, Hungary
[8]First Department of Medicine, University of Pécs Medical School, Pécs, Hungary
[9]Division of Gastroenterology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
[10]Department of Gastroenterology, Hospital Universitario Central de Asturias, Oviedo, Spain
[11]Department of Surgery, Iuliu Hatieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania
[12]Department of Surgery, Professor Doutor Fernando Fonseca Hospital, Amadora, Portugal
[13]Department of Gastroenterology, University Clinical Hospital of Valladolid, Valladolid, Spain
[14]Department of Gastroenterology, Lozano Blesa University Clinic Hospital, Aragon Health Research Institute (IIS Aragón), Zaragoza, Spain
[15]Department of Surgery, Hospital Regional de Alta Especialidad del Bajío, León, Mexico
[16]General Surgery Department 1, Tondela-Viseu Hospital Centre, Viseu, Portugal
[17]Department of General Surgery, Emergency Hospital of Bucharest, Carol Davila University of Medicine and Pharmacy, Bucharest, Romania
[18]Gastroenterology Department, Hospital Universitario Miguel Servet, Zaragoza, Spain
[19]Medical Department II (Gastroenterology, Hepatology, Infectious Diseases, Pulmonology), University of Leipzig Medical Center, Leipzig, Germany
[20]Department of Surgical Disciplines, Immanuel Kant Baltic Federal University, Regional Clinical Hospital, Kaliningrad, Russian Federation
[21]Department of Gastroenterology, Complejo Hospitalario de Navarra, Pamplona, Spain
[22]Gastroenterology Department, Hospital Universitario San Ignacio/Universidad Javeriana, Bogota, Colombia
[23]Department of Gastroenterology, Hospital Clínico Universitario de Valencia. Universidad de Valencia, Valencia, Spain
[24]Hospital Provincial-Pla Health Center, Alicante-Sant Joan Health District, Alicante, Spain

**Twitter** Enrique de-Madaria @DeMadaria and Claudia Sánchez-Marin @DraClauS

**Acknowledgements** We are especially grateful to the international team of leading pancreatologists who helped to design the PAN-PROMISE scale: Pramod K Garg (New Delhi, India), Markus Lerch (Greifswald, Germany), Félix Lluís (Alicante, Spain), Juan Martínez (Alicante, Spain), Georgios Papachristou (Pittsburgh, USA), Jonas Rosendahl (Leipzig, Germany), Hakan Senturk (Istanbul, Turkey), Vikesh K Singh (Baltimore, USA) and Bechien Wu (Los Angeles, USA). We would also like to thank the professionals from the Alicante region who participated in the nominal group during the design of the PAN-PROMISE scale: Alicia Brotons, Isabel de Castro, Marta Garmendia, Macarena García, Rolando Illescas, Belén Martínez, Juan Martínez and Laura Sempere. Thanks to those colleagues who helped us with the translations of the scale but didn't participate in the cohort study: Jorge Machicado, Robert Moran, Raffaele Pezzilli, Livia Archibugi, Ahmad Uzair Qureshi, Sun Xiaotian, Vinciane Rebours, Enrique Pérez-Cuadrado Robles, Arda Isik, Erkan Mert, Vladimir Khokha and Mariia Kiriukova.

**Collaborators** Chuklin Serhii; Shershen Gennadiy; Tudor Cerciu; Sergiu Matei; Víctor Jiménez-Beltrán; Rita V. Martins; Rosa Matias; Antonio Gomes; Daniel Abad-Baroja; José P. Pinto; Ana C. Afonso; Cezar Ciubotaru; Bogdan Stoica; Vanesa Bernal; Eduardo Bajador-Andreu; Carlos Marra-López; Juliana Mantilla-Carreño; Claudia Zuluaga-Serna; Anita Gasiorowska; Natalia Fabisiak; László Czakó; Dóra Illés; Ainhoa Abando-Zurimendi; María Fraile; Young Koog Cheon; Konstantinos Triantafyllou; Polychronis Gatos-Gatopoulos; Rubén Piqué; Iván Guerra; Juan A. Rodríguez-Oballe; Beatriz Romero-Mosquera; Giulia De Marchi; Yonko Georgiev; Aurora Mula.

**Contributors** All authors were involved in data acquisition in phase I and/or II (except SSV), interpretation of the data and critical revision of the manuscript. Study concept and design: Ed-M and JJM. Primary analysis of the data: JJM, Ed-M, CS-M and IC. External review and assessment: SSV. Initial draft of the manuscript: JJM, Ed-M, PH, CS-M and SSV. Study supervision: CS-M. All authors have approved the final version of the manuscript.

**Funding** This study received no funding. The Spanish Association of Gastroenterology (AEG) provided access to a tool to develop the electronic case report form (REDCap) free of charge, with the sole aim of promoting independent investigator-driven research.

**Competing interests** None declared.

**Patient and public involvement** Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

**Patient consent for publication** Not required.

**Ethics approval** The study was initially approved on 09/01/2017 by the Ethic Committee of Clinical Research of the Alicante University General Hospital (Alicante, Spain), project identification code PI2016/69, and later by the Institutional Review Boards of all collaborating centres.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available on reasonable request. Study data are available to our collaborators to perform collateral studies.

**Author note** In loving memory of João Conde.

**ORCID iDs**
Enrique de-Madaria http://orcid.org/0000-0002-2412-9541
Serge Chooklin http://orcid.org/0000-0002-3503-8450
Peter Hegyi http://orcid.org/0000-0003-0399-7259
Ayesha Kamal http://orcid.org/0000-0001-5355-9199
Marcus Hollenbach http://orcid.org/0000-0002-2654-3164

## REFERENCES

1 Peery AF, Crockett SD, Murphy CC, *et al*. Burden and cost of gastrointestinal, liver, and pancreatic diseases in the United States: update 2018. *Gastroenterology* 2019;156:254-272.e11.
2 Sternby H, Bolado F, Canaval-Zuleta HJ, *et al*. Determinants of severity in acute pancreatitis: a nation-wide multicenter prospective cohort study. *Ann Surg* 2019;270:348–55.
3 Banks PA, Bollen TL, Dervenis C, *et al*. Classification of acute pancreatitis--2012: revision of the Atlanta classification and definitions by international consensus. *Gut* 2013;62:102–11.
4 Párniczky A, Kui B, Szentesi A, *et al*. Prospective, multicentre, nationwide clinical data from 600 cases of acute pancreatitis. *PLoS One* 2016;11:e0165309.
5 Vege SS, DiMagno MJ, Forsmark CE, *et al*. Initial medical treatment of acute pancreatitis: American gastroenterological association Institute technical review. *Gastroenterology* 2018;154:1103–39.
6 Forsmark CE, Vege SS, Wilcox CM. Acute pancreatitis. *N Engl J Med* 2016;375:1972–81.
7 Wu BU, Batech M, Quezada M, *et al*. Dynamic measurement of disease activity in acute pancreatitis: the pancreatitis activity scoring system. *Am J Gastroenterol* 2017;112:1144–52.
8 Nelson E*et al*. Functional health status levels of primary care patients. *JAMA* 1983;249:3331–8.
9 Lloyd H, Wheat H, Horrell J, *et al*. Patient-Reported measures for Person-Centered coordinated care: a comparative domain map and web-based compendium for supporting policy development and implementation. *J Med Internet Res* 2018;20:e54.
10 US Food and Drug Administration CfDEaR, Center for Biologics Evaluation and Research, Center for Devices and Radiological Health. Guidance for industry patient-reported outcome measures: use in medical product development to support labeling claims, 2009. Available: http://purl.access.gpo.gov/GPO/LPS1134132019
11 Fitzpatrick R, Davey C, Buxton MJ, *et al*. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess* 1998;2:1–74.
12 Nelson EC, Eftimovska E, Lind C, *et al*. Patient reported outcome measures in practice. *BMJ* 2015;350:g7818.
13 Dawson J, Doll H, Fitzpatrick R, *et al*. The routine use of patient reported outcome measures in healthcare settings. *BMJ* 2010;340:c186.
14 Black N, Varaganum M, Hutchings A. Relationship between patient reported experience (PREMs) and patient reported outcomes (PROMs) in elective surgery. *BMJ Qual Saf* 2014;23:534–42.
15 Boyce MB, Browne JP, Greenhalgh J. The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research. *BMJ Qual Saf* 2014;23:508–18.
16 Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. *BMJ Open* 2013;3:e001570.
17 Terwee CB, Prinsen CAC, Chiarotto A, *et al*. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 2018;27:1159–70.
18 Mokkink LB, Terwee CB, Patrick DL, *et al*. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–45.
19 World Medical A. World Medical association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 2013;310:2191–4.
20 Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. *J Eval Clin Pract* 2006;12:559–68.
21 Organization WH. Process of translation and adaptation of instruments www.who.int: World Health organization. Available: https://www.who.int/substance_abuse/research_tools/translation/en/2019
22 Karnofsky DAB JH. The clinical evaluation of chemotherapeutic agents in cancer. In: MacLeod CM, ed. *Evaluation of chemotherapeutic agents*. New York: Columbia University Press. In Press, 1949: 191–205.
23 Aaronson NK, Ahmedzai S, Bergman B, *et al*. The European organization for research and treatment of cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993;85:365–76.
24 Harris PA, Taylor R, Thielke R, *et al*. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377–81.
25 Gluud C, Krogsgaard K. Would you trust a surrogate respondent? *The Lancet* 1997;349:665–6.
26 Freemantle N, Calvert M, Wood J, *et al*. Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA* 2003;289:2554–9.
27 Arai S, Stotts N, Puntillo K. Thirst in critically ill patients: from physiology to sensation. *Am J Crit Care* 2013;22:328–35.
28 de-Madaria E, Banks PA, Moya-Hoyo N, *et al*. Early factors associated with fluid sequestration and outcomes of patients with acute pancreatitis. *Clin Gastroenterol Hepatol* 2014;12:997–1002.
29 Huang W, de la Iglesia-García D, Baston-Rey I, *et al*. Exocrine pancreatic insufficiency following acute pancreatitis: systematic review and meta-analysis. *Dig Dis Sci* 2019;64:1985–2005.