

## Original Article

**Cite this article:** Jones, A. S.K., Seaton, N., Brown, A., Jenkinson, E., Carroll, S., Dietz, K. C., Hudson, J. L., Wroe, A., & Moss-Morris, R. (2025). The Illness-Related Distress Scale: development and psychometric evaluation of a new transdiagnostic measure. *Psychological Medicine*, 55, e122, 1–13  
<https://doi.org/10.1017/S003329172500090X>

Received: 27 November 2024

Revised: 14 March 2025

Accepted: 24 March 2025








**Corresponding author:**

Rona Moss-Morris;

Email: [rona.moss-morris@kcl.ac.uk](mailto:rona.moss-morris@kcl.ac.uk)

A.S.K.J. and N.S. are joint first authors and co-primary authors.

# The Illness-Related Distress Scale: development and psychometric evaluation of a new transdiagnostic measure

Annie S. K. Jones , Natasha Seaton , Ashley Brown, Emma Jenkinson , Susan Carroll , Kristina C. Dietz, Joanna L. Hudson , Abigail Wroe  and Rona Moss-Morris 

Health Psychology Section, Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

**Abstract**

**Background.** Individuals with long-term physical health conditions (LTCs) experience higher rates of depression and anxiety. Conventional self-report measures do not distinguish distress related to LTCs from primary mental health disorders. This difference is important as treatment protocols differ. We developed a transdiagnostic self-report measure of illness-related distress, applicable across LTCs.

**Methods.** The new Illness-Related Distress (IRD) scale was developed through thematic coding of interviews, systematic literature search, think-aloud interviews with patients and healthcare providers, and expert-consensus meetings. An internet sample ( $n = 1,398$ ) of UK-based individuals with LTCs completed the IRD scale for psychometric analysis. We randomly split the sample (1:1) to conduct: (1) an exploratory factor analysis (EFA;  $n = 698$ ) for item reduction, and (2) iterative confirmatory factor analysis (CFA;  $n = 700$ ) and exploratory structural equation modeling (ESEM). Here, further item reduction took place to generate a final version. Measurement invariance, internal consistency, convergent, test–retest reliability, and clinical cut-points were assessed.

**Results.** EFA suggested a 2-factor structure for the IRD scale, subsequently confirmed by iteratively comparing unidimensional, lower order, and bifactor CFAs and ESEMs. A lower-order correlated 2-factor CFA model (two 7-item subscales: *intrapersonal* distress and *interpersonal* distress) was favored and was structurally invariant for gender. Subscales demonstrated excellent internal consistency, very good test–retest reliability, and good convergent validity. Clinical cut points were identified (*intrapersonal* = 15, *interpersonal* = 12).

**Conclusion.** The IRD scale is the first measure that captures transdiagnostic distress. It may aid assessment within clinical practice and research related to psychological adjustment and distress in LTCs.

**Introduction**

Approximately, 50% of UK adults live with at least one long-term health condition (LTC) (Office for National Statistics, 2022). Adjusting to an LTC can be challenging given the burden of symptoms and treatments, physical disability, loss of independence, and reduced quality of life. Unsurprisingly, people living with LTCs are 2–3 times more likely to have anxiety and depression than those without (McDaid, Knapp, Fossey, & Galea, 2012). Yet many of these patients do not have access to psychological support for their illness (Diabetes UK, 2019; Ellison, Gask, Bakerly, & Roberts, 2012; IBD UK, 2021; Ponzio, Tacchino, Zaratin, Vaccaro, & Battaglia, 2015; Schwarz, Schmidt, Bobek, & Ladurner, 2022). When offered treatment, this is often targeted to depression and anxiety as a primary mental health condition rather than the unique LTC stressors that can lead to illness-related distress.

The transdiagnostic model of adjustment in LTCs (TMA-LTC) (Carroll, Moon, Hudson, Hulme, & Moss-Morris, 2022) suggests that poor psychological adjustment to an LTC or LTC-related distress results in part from unique illness-specific stressors (e.g. stigma, symptom and treatment management, uncertainty about the future) which are distinct from primary mental health risk factors such as low self-esteem or global hopelessness. Though different LTCs have a specific set of stressors and self-management demands, there are core transdiagnostic mechanisms underlying psychological adjustment and LTC distress. Helping people manage these illness stressors should be central to psychological therapy for people with LTC-related distress. Being able to distinguish LTC distress from a primary mental health disorder is an important first step in ensuring LTC patients get the correct psychological support (Carroll, Moss-Morris, Hulme, & Hudson, 2021).

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use and/or adaptation of the article.



CAMBRIDGE  
UNIVERSITY PRESS

The Patient Health Questionnaire (PHQ-9) (Kroenke, Spitzer, & Williams, 2001) and Generalized Anxiety Disorder Questionnaire (GAD-7) (Spitzer, Kroenke, & Williams, 2006) are commonly used to screen patients for possible mental health disorders. Whilst these measures have excellent psychometric properties and well-validated cut points for clinical caseness, they have limitations in screening for LTC distress. First, negative emotions associated with poor LTC adjustment extend beyond anxiety and depression, including feelings of anger, guilt, embarrassment, and shame (Ayers & Steptoe, 2007; Browne, Ventura, Mosely, & Speight, 2013; Kreider, 2017). Second, patients distressed by their illness may score subthreshold on traditional measures of anxiety and depression (Geraghty & Esmail, 2016; Katon & Roy-Byrne, 1991), due to these inadequately capturing LTC distress. Third, relating adjustment to diagnostic levels of anxiety and depression may unnecessarily pathologize the negative emotions resulting from objectively challenging illness-related stressors (Hudson & Moss-Morris, 2019). Finally, some anxiety/depression symptoms are common symptoms of LTCs (e.g. fatigue, sleep disturbances), obscuring the unique distress experienced due to poor adjustment.

Therefore, there is a need to measure LTC-related distress to aid clinical decision-making. In LTC care, Distress Thermometers alongside Problem Lists are sometimes used; however, these have some important psychometric limitations. Distress Thermometers and other single-item measures inadequately capture complex psychological constructs (Allen, Iliescu, & Greiff, 2022; Cuvillier, Léger, & Sénécal, 2021; Stewart-Knight, Parry, Abey, & Seymour, 2012). Problem Lists have clinical utility in identifying sources of distress but they do not measure the severity of distress or allow comparisons across conditions. Generic psychological distress measures such as the Kessler K-10 scale (Kessler *et al.*, 2002) and Patient Health Questionnaire Anxiety and Depression Scale (PHQ-ADS) (Kroenke *et al.*, 2016) effectively assess the severity of distress; however, they do not differentiate whether the distress is related to an individual's LTC or an unrelated mental health disorder or other non-LTC life stressors. Conversely, illness-specific distress measures exist for some LTCs (e.g. inflammatory bowel disease [IBD]; Dibley *et al.*, 2018 and diabetes; Fisher, Glasgow, Mullan, Skaff, & Polonsky, 2008). However, there is no transdiagnostic measure of illness-specific distress that can be used across various LTC populations. A transdiagnostic measure has greater utility in primary care or mental health services that are not specialized to particular LTCs while minimizing administrative burden. For instance, in the UK Talking Therapy services, healthcare professionals report having low confidence in determining whether an LTC treatment is appropriate and wanting additional tools and skills to assess and treat these patients (Carroll *et al.*, 2021). Therefore, a transdiagnostic measure of IRD could be used alongside more traditional measures of distress, anxiety, and depression to signal whether a primary mental health or LTC adjustment protocol should be used (Carroll *et al.*, 2022; Jenkinson, Hudson, Moss-Morris, & Hackett, *in prep.*). Moreover, as multimorbidity is increasingly common, estimated to affect over 50% of UK and US populations (Fleetwood *et al.*, 2025; Head *et al.*, 2021; Knies & Kumari, 2022; Mossadeghi *et al.*, 2023), and appears to confer additional risk of distress (Fleetwood *et al.*, 2025; Read, Sharpe, Modini, & Dear, 2017), a transdiagnostic measure would be better placed to capture the additive impact of multiple health concerns. Furthermore, it would cater for rarer conditions and would allow comparison of LTC distress across conditions in both clinical and research settings.

The primary aims of the current study were to develop a novel, concise transdiagnostic measure of illness-related distress (IRD) with good face validity and to assess the factor structure and the minimal number of best-fit items, convergent validity, internal consistency, and test-retest reliability of the scale. A secondary aim was to explore clinical cut points of the scale using Receiver Operating Characteristic (ROC) analyses to guide clinical decision-making and treatment assessment.

## Methods

The study was registered on [clinicaltrials.gov](https://clinicaltrials.gov) (NCT06072287). Ethics approval was obtained from the King's College London Health Faculties Research Ethics Subcommittee on the July 13, 2023 (HR/DP-22/23-36320). All participants provided informed consent.

### Procedures and recruitment

Eligibility criteria were: self-reporting a LTC; being UK-based; being aged  $\geq 18$  years; having an email address; and having English proficiency. Participants were excluded if they only reported psychological or mental disorders.

We conducted convenience sampling via social media and charity website advertisements (Supplement 1 of the [Supplementary Material](#)). Links directed participants to the information sheet, followed by eligibility screening, consent, and the baseline questionnaire. To assess test-retest reliability, 1 week later, respondents were emailed a link to complete a follow-up questionnaire (IRD scale only).

## Measures

### The Illness-Related Distress (IRD) Scale

Several pieces of formative research summarized in [Table 1](#) shaped the initial selection of items for the IRD scale with a focus on ensuring good face validity of the items.

A preliminary 28-item scale was tested in the current study (Supplement 3 of the [Supplementary Material](#)). Respondents reported the frequency with which they had experienced each item during the past 2 weeks. Items were scored on a five-point Likert-type response scale from 0 ('Never') to 4 ('Always'); five items were reverse scored.

A slider item was included as a validity check, whereby participants rated the source of their distress, ranging from 'entirely due to other life stressors' (0%) to 'entirely due to their LTC' (100%). Respondents could select N/A if they did not feel distressed.

### Demographics

At baseline, respondents provided their age, gender, ethnicity, level of education, employment status, and LTC diagnoses. LTC response options were determined via gold-standard studies in LTCs and the National Institute for Health and Care Excellence (NICE) guidelines (Coulter *et al.*, 2015; NICE, 2024).

### Assessment of validity

Self-report measures to assess the validity of the IRD scale were informed by the COSMIN Taxonomy of Measurement Properties (Mokkink *et al.*, 2010). Measures were selected to maximize

**Table 1.** Summary of methods used to develop the initial 28-item pool of the IRD scale

Initial item generation	<p><i>Qualitative interviews</i></p> <ul style="list-style-type: none"> <li>- In a previous study (Hulme et al., 2021; Picariello et al., 2024), 30 participants with a range of LTCs took part in interviews (kidney disease, MS, psoriasis, and IBD)</li> <li>- Participants were asked open-ended questions about their perceptions of the relationship between their illness and mood</li> <li>- Transcripts were inductively coded (ASKJ) to identify key topics and stressors related to IRD</li> <li>- Patients were asked about distress, though this word did not appear to be particularly meaningful. Patients generally identified with a broader spectrum of emotions, including worry, anger, frustration, guilt, and embarrassment</li> </ul> <p><i>Systematic literature search</i></p> <ul style="list-style-type: none"> <li>- A systematic search of EMBASE and Medline (conducted on November 21, 2022) identified existing disease-specific tools measuring IRD (search terms in Supplement 2 of the Supplementary Material).</li> <li>- 7,344 were identified (<math>n = 10,020</math> before deduplication) and were double-screened by NS and EJ.</li> <li>- 38 unique measurement tools were identified. The content and style were reviewed to guide the initial IRD scale items.</li> <li>- Two scales were selected to be included in subsequent validity testing: The IBD Distress Scale (IBDDS) (Dibley et al., 2018) and the Diabetes Distress Scale (DDS) (Fisher et al., 2008)</li> </ul> <p><i>Theory</i></p> <ul style="list-style-type: none"> <li>- The Transdiagnostic Model of Adjustment to LTC (Carroll et al., 2022) helped with conceptualization of illness-specific stressor items</li> </ul>
Refining draft	<p><i>Drafting</i></p> <ul style="list-style-type: none"> <li>- Based on the initial item generation stage, four authors (AJ, EJ, NS, SC), with expertise in clinical research and clinical practice in LTC groups, drafted an initial item list.</li> <li>- Items were reviewed by a wider group of authors (ASKJ, NS, EJ, JH, AW, SC, RMM), with additional expertise in delivering and developing therapies for treating IRD and scale development.</li> <li>- Over three consensus meetings, the initial list was refined to list of 32 items, a slider scale, and an open-ended item asking respondents to list the LTC(s) they were thinking of most whilst answering.</li> <li>- A 2-week timeframe was chosen to align with the PHQ-9, GAD-7, and PHQ-ADS scales. The 2-week timeframe allows for reliable assessment of recent symptoms over a representative period of time. It enables identification of trends, and changes in mood that might not be apparent if looking at a shorter timeframe, while still being manageable for patients to recall accurately.</li> </ul>
Feedback from people with lived experience	<p><i>Think-aloud interviews with people living with LTCs (<math>n = 14</math>)</i></p> <ul style="list-style-type: none"> <li>- To ensure face validity of the items, people living with a variety of LTCs were invited to provide feedback on the refined list of 32 items. 'Think aloud' methods (Charters, 2003) were used, with participants speaking their thoughts as they read through the items. This allowed the researchers to capture how respondents might interpret each of the items.</li> <li>- Changes were made to items to remove ambiguity, and increase relevance and readability. Four items were deemed not helpful or useful and so were excluded, leaving 28 items.</li> </ul> <p><i>Feedback from HCPs</i></p> <ul style="list-style-type: none"> <li>- The questionnaire was reviewed over email by seven HCPs (three clinical psychologists, two health psychologists, one clinical health psychologist, and one specialized counseling psychologist)</li> <li>- HCPs gave feedback on the relevance of the scale and coverage of key concepts. Changes were made to improve wording or give additional examples.</li> </ul>

Note: HCPs, healthcare professionals; IBD, inflammatory bowel disease; LTC, long-term condition; MS, multiple sclerosis; PPI, patient and public involvement; TMA-LTC, Transdiagnostic Model of Adjustment in Long-Term Conditions.

relevance to our transdiagnostic populations while minimizing participant burden.

To assess convergent validity, we measured:

- *Psychological distress, depression, and anxiety:* The PHQ Anxiety and Depression Scale (PHQ-ADS) (Kroenke et al., 2016) combines the Patient Health Questionnaire-8 (PHQ-8) (Kroenke et al., 2009) for depression and the Generalized Anxiety Disorder-7 (GAD-7) (Spitzer et al., 2006) scale for anxiety to create an overall measure of psychological distress. All measures were responded to on a four-point Likert scale (0–3) and utilized sum scores. Higher scores indicate greater levels of distress/depression/anxiety. Here, Cronbach  $\alpha_{PHQ-ADS} = 0.92$ ;  $\alpha_{PHQ} = 0.85$ ;  $\alpha_{GAD} = 0.91$ .
- *Illness-specific distress:* The Diabetes Distress Scale (DDS) is a 17-item diabetes-related distress questionnaire (Fisher et al., 2008),  $\alpha_{DDS\text{baseline}} = 0.95$ . The 28-item IBD Distress Scale (IBDDS), measures distress in IBD (Dibley et al., 2018)  $\alpha_{IBDDS\text{baseline}} = 0.93$ ). Higher total scores on each measure indicate increased distress. Only participants who identified as having diabetes or IBD completed the DDS or IBDDS, respectively.

- *Functional impairment:* The Work and Social Adjustment Scale (WSAS) (Mundt, Marks, Shear, & Greist, 2002) measures overall impairment in everyday life using five items. Higher total summed scores indicate greater impairments in functioning,  $\alpha_{WSAS} = .88$ .
- *Cognitive and Behavioral Responses to Symptoms:* The Cognitive and Behavioral Responses to Symptoms Questionnaire (CBRQ) (Picariello, Chilcot, Chalder, Herdman, & Moss-Morris, 2023) has 40-items with seven subscales: five cognitive (Fear Avoidance, Catastrophizing, Damage Beliefs, Embarrassment Avoidance, and Symptom Focusing) and two behavioral (All-or-Nothing Behavior, Avoidance/Resting behavior) subscales. Higher summed scores indicate a stronger presence of the specific cognitive/behavioral response;  $\alpha_{CBRQ\text{ subscale range}} = 0.80\text{--}0.91$ .

### Readability

Readability was assessed with The Flesch Reading Ease score, providing a score out of 100 and the reading age at which the material is appropriate.

## Statistical analyses

Data were analyzed between March 8, 2024 and March 14, 2025 (available in online repository: <https://osf.io/gnwe6/>).

### Step 1: Characteristics of samples

The sample was randomly split into two groups using a random number generator (Microsoft Excel v2402) to allow for an initial exploratory factor analysis (EFA) in one sample followed by a confirmatory factor analysis (CFA) in the other. The rationale for using both methods is described in the next sections. Descriptive statistics were performed in STATA v18.0 for the total sample and the subgroups.

### Step 2: Factor analysis

Factor analysis was used to assess the factor structure and best-fit items of the IRD. Unless otherwise specified, steps 2.1 (EFA), 2.2 (CFA/ESEM), and 4 (Invariance testing) of the analysis were conducted using MPlus V 7.4. A maximum likelihood with robust standard errors (MLR) was used to treat missing data and account for non-normality. Given the five response categories, the ordinal data from each item was treated as continuous (Rhemtulla, Brosseau-Liard, & Savalei, 2012).

For best practice, exploratory and confirmatory models should be conducted in different samples to reduce bias and risk of over-fitting. The EFA was conducted in Sample 1 ( $n = 698$ ). All model fitting (step 2.2) was conducted with Sample 2 ( $n = 700$ ). The sample-to-item ratio exceeded the recommended 10:1 (Costello & Osborne, 2019) (Sample 1 ratio: 24.9:1; Sample 2 ratio: 25:1).

#### Step 2.1: Exploratory factor analysis (EFA)

EFA was used to (1) reduce the item pool and (2) determine the general factor structure (Rhemtulla et al., 2012). Factors were examined for item loadings and eigenvalues. Factors with eigenvalues  $\geq 1$  and a minimum of three items per factor were retained (Costello & Osborne, 2019). This approach, although seen as excessively liberal (Cliff, 1988; Horn, 1965) was chosen due to the exploratory nature of this scale development. Initially, items with primary factor loadings of  $\leq 0.4$  were eliminated, and  $\leq 0.45$  were investigated further. Items were removed if cross-loadings between primary and secondary factors were  $< 0.15$ . Factors in the final EFA, selected based on eigenvalues, root mean square error of approximation estimation (RMSEA; values of 0.01, 0.05, and 0.08 indicate excellent, good, and mediocre fit, respectively) (Xia & Yang, 2019), and factor loadings, in conjunction with theory and previous evidence, were used in subsequent CFA and ESEM analyses. Importantly, we did not let the  $\lambda \geq 1$  dictate our final model and ran a parallel analysis to further guide our decisions. Face and construct validity were prioritized rather than pre-emptively restricting models in our item/factor reduction steps.

#### Step 2.2: Model fitting

ESEM is an integrative approach that balances the strictness of CFA and the adaptability of EFA (Marsh et al., 2014). CFAs often use overly restrictive models where cross-loadings between items and non-target factors are fixed at zero. Consequently, CFAs may not always yield a good model fit or assist in the theoretical interpretation of multidimensional constructs, particularly those with multiple factors (Brown, Barker, & Rahman, 2022; Dicke et al., 2018; Marsh, Hau, & Grayson, 2005; Morin, Arens, Tran, & Caci, 2016). This is common with psychological constructs; items rarely perfectly define a construct due to potential association with similar

constructs or sub-dimensions. ESEM incorporates an EFA measurement model using target rotation, allowing for confirmatory use by specifying a priori cross-loadings (Asparouhov & Muthén, 2009; Morin, Myers, & Lee, 2020; Morin et al., 2016). Cross-loadings for non-targeted items are set close to zero, avoiding unnecessary restrictions as in CFA. We used an analytic framework to systematically compare CFA and ESEM hierarchical models (Morin et al., 2020).

Three CFA models were investigated: (1) unidimensional, (2) correlated factors (specified by the final EFA), and (3) bifactor model with one general factor and specific factors (orthogonal). Two ESEM models were assessed: (1) correlated lower order factors, and (2) bifactor model. Superiority of models would be decided by: (1) better model fit, (2) smaller factor correlations, (3) smaller cross-loadings, and (4) well-defined factors (Morin et al., 2020; Morin et al., 2016). Bifactor model superiority would be confirmed if there was (1) an improved fit in comparison to lower-order correlated factor models and (2) well-defined general and specific factors.

Absolute model fit was assessed with the  $\chi^2$  goodness of fit statistic (non-significant ( $p > .05$ ) values indicating good fit) and the standardized root mean square residual (SRMR, with values  $< 0.05$  indicating good and  $< 0.08$  indicating acceptable fit), respectively, for both indices. Both indices were deemed necessary as the  $\chi^2$  goodness of fit statistic typically rejects models with large sample sizes (Hooper, 2008).

The relative fit was using Hu and Bentler (1999) recommendations of the two-fit criterion, that the comparative fit index (CFI) and Tucker-Lewis index (TLI) should be  $> .95$  and RMSEA and SRMR should be  $< .06$  to minimize Type 1 and Type 2 error rates. The Akaike Information Criteria (Akaike, 1987) and the Bayesian Information Criteria (BIC; Schwarz, 1978) were also used to assess relative fit, where lower values indicate improved fit.

The Omega ( $\omega$ ) program was used (Watkins, 2013) to estimate fit indices for the final bifactor models, as Cronbach's  $\alpha$  is limited in the assumption of equal factor loadings across all constructs for each indicator (Dunn, Baguley, & Brunsden, 2014).

Several  $\omega$  coefficient variants were used to assess if there was sufficient variance accounted for by both general and specific factors to justify the selection of a hierarchical bifactor model (Rodriguez, Reise, & Haviland, 2016). We also assessed construct replicability (H), explained common variance (ECV), and the percentage of uncontaminated correlations (PUC).

### Step 3: Creating a shortened clinical IRD scale

The study aimed to create a brief questionnaire for use in clinical settings. Items were considered for removal if they: were thematically similar; had primary factor loadings  $\leq 0.55$  (in CFA analysis); had a cross-loading  $\leq 0.20$  (in EFA analysis); or correlated highly ( $r \geq 0.6$  in CFA analysis). Step 2.2 was repeated for the final clinical version.

### Step 4: Measurement invariance

Measurement invariance was assessed across gender in the final best-fitting model for the initial and final clinical versions. We assessed (1) configural (factor structure), (2) metric (factor loadings), and (3) scalar (item intercepts) invariance to see if these patterns were stable across groups (Morin et al., 2016).

The models for each invariance sub-group were compared with the models that arose from Step 2 and 3. The same indices to assess the goodness of fit in Step 2 were used to assess invariance models.



Since  $\chi^2$  is sensitive to sample size, we also considered changes in CFI ( $\Delta$ CFI), RMSEA ( $\Delta$ RMSEA), and SRMR ( $\Delta$ SRMR) for invariance decisions. Significant levels of invariance were decided a priori, reflecting similar analyses (Brown et al., 2022). Cut-offs were: CFI decrement ( $\Delta$ CFI)  $<0.010$ ; RMSEA change ( $\Delta$ RMSEA)  $<0.015$  (Chen, 2007);  $\Delta$ SRMR  $<0.030$  for configural and metric invariance and  $<0.010$  for scalar invariance.

#### Step 5: Validity and reliability testing

Additional psychometric testing was conducted in STATA V18. Total subscale scores were calculated using weighted factor scores. Due to the expected clinical utility of the scale, unweighted totals (sum of items, irrespective of factor loading) were also calculated.

Criterion validity testing was conducted on both initial and final clinical questionnaire versions (weighted and unweighted scores). Pearson's bivariate correlations assessed test-retest reliability and convergent validity of the latent factors from the final best-fit model (Step 2.2). Effect sizes were interpreted using standard cut-offs (Cohen, 2013). Convergent validity was supported through moderate correlations (usually,  $|r| = 0.4$ – $0.6$ ), whereas, strong correlations indicate construct overlap or multicollinearity. Internal consistency was assessed using both Cronbach's  $\alpha$  and McDonald's  $\omega$  estimates (target values  $>0.70$ ) (Bland & Altman, 1997; Nunnally & Bernstein, 1978).

#### Step 6: Receiver operating characteristic (ROC) analyses

ROC analyses with boot-strapped 95% Confidence Intervals ( $CI_{95}$ , based on 10,000 iterations) were used to determine optimal cut-points, with equal emphasis on sensitivity and specificity (Youden, 1950). In this context, sensitivity is important as illness-related distress can result in suicidal ideation and behavior. Specificity is important, as clinical support is expensive and the capacity of health services is limited. The slider item served as the true class/gold standard for caseness, with participants attributing at least 50% of their psychological distress to their LTC considered as a case. ROC analyses were conducted in R version 4.1.2 (2021-11-01), using the pROC package (Robin et al., 2011).

## Results

### Characteristics of study samples

We recruited participants from June 28, 2023 to January 22, 2024. There were 2,114 entries of the baseline questionnaire; however, 474 were removed as they were suspected to be automated 'bot' submissions. Responses were deemed at a high likelihood of automation if they were excessively similar (e.g. multiple submissions had identical response patterns), had suspicious responses (e.g. providing numerical postcodes when the study is UK-based), faster than expected response times, and/or non-UK IP addresses. This left 1,640 authentic entries of the baseline questionnaire. Of these, 242 entries were removed leaving a total of 1,398 in the baseline sample (participant flow in Supplement 4 of the Supplementary Material).

Follow-up was completed by 1,240 participants (88.7% response rate), with 1,171 completing their follow-up questionnaire between 6 and 48 days after baseline ( $M = 11.51$ ,  $SD = 7.71$ ). Table 2 shows the demographics of respondents and the most common illnesses reported. Supplement 5 of the Supplementary Material shows a full list of the 198 'Other' LTCs reported.

### Exploratory factor analysis

The initial EFA with 28 items yielded four factors with eigenvalues  $>1$ . Additionally, we performed a parallel analysis (O'Connor, 2000) with 5,000 random samples, as this is considered more statistically robust. This similarly suggested four factors be retained ( $\lambda_1$ : 12.95 vs.  $\lambda_{1,random sample} = 1.39$ ;  $\lambda_2$ : 5.52 vs.  $\lambda_{2,random sample} = 1.33$ ;  $\lambda_3$ : 2.36 vs.  $\lambda_{3,random sample} = 1.29$ ;  $\lambda_4$ : 1.49 vs.  $\lambda_{4,random sample} = 1.26$ ). Variance explained by each factor was 46.2%, 19.70%, 8.30%, and 5.3%, respectively. One factor was removed as it only contained two items, and another was investigated as it only contained three items.

In subsequent EFAs, items were removed if primary loadings were  $<|.40|$  or if cross-loadings were within 0.15 of primary loadings. Consequently, the final EFA utilized 23 of the original 28 items and had three factors with eigenvalues  $>1$  ( $\lambda_1$ : 11.48;  $\lambda_2$ : 1.34;  $\lambda_3$ : 1.04); parallel analysis suggested two factors ( $\lambda_{1,random sample} = 1.34$ ;  $\lambda_{2,random sample} = 1.28$ ;  $\lambda_{3,random sample} = 1.24$ ). This, taken together with factor inspection, suggested the three-factor model was inappropriate as the third factor contained only reverse-coded items, a common phenomenon in scale development and psychometrics (Salazar, 2015). Thus, the two-factor EFA model was selected:  $\chi^2(208) = 932.32$ ,  $p < .001$ , RMSEA = .086, 90% CI [.081, .090], RMSR = .052. The factors were named (1) *intrapersonal* distress (14 items; 48.29%) and (2) *interpersonal* distress (9 items; 19.75%), ( $r = 0.811$ ,  $p < 0.001$ ) (Primary rotated loadings in Supplement 7 of the Supplementary Material).

### Model fitting (CFA and ESEM)

For the *IRD<sub>initial</sub>* the lower order CFA model was most appropriate. Factor loadings and bifactor fit indices indicated that a 'g' factor was inappropriate. The lower order CFA was preferred over the lower order ESEM given that the former was more parsimonious with only minimal differences in fit and factor loadings compared with the latter (Supplement 8 of the Supplementary Material). We subsequently removed nine items as described in Step 3 to create the *IRD<sub>final</sub>* with seven items defining each factor. Factors remained strongly and positively correlated in this model ( $r = 0.774$ ,  $p < 0.001$ ). All testing (Steps 4–6) was performed on both *IRD<sub>initial</sub>* and *IRD<sub>final</sub>*. For brevity *IRD<sub>initial</sub>* is presented in Supplement 8 of the Supplementary Material; all results presented below pertain to *IRD<sub>final</sub>*.

Amongst CFA models, the bifactor model demonstrated the lowest value for  $\chi^2$ , RMSEA, SRMR, AIC, and BIC and the highest CFI and TLI (Table 2), but poorer model fit with respect to standardized item loadings on each factor (Supplement 9 of the Supplementary Material). We rejected a unidimensional model as, although factor loadings were adequate, the two-factor model had good-excellent fit and demonstrated face validity (Supplement 9 of the Supplementary Material).

In ESEM models, the bifactor model demonstrated superior fit indices (Table 3). Items loaded well onto a general factor in the bifactor model, however loadings onto both specific factors were inadequate/inconsistent, suggesting a non-hierarchical model was more appropriate (Supplement 10 of the Supplementary Material). In additional calculations, both bifactor models supported the rejection of hierarchical CFA and ESEM models (Supplement 11 of the Supplementary Material). Specific factors (labeled *intrapersonal* distress and *interpersonal* distress) demonstrated high internal reliability (acceptable cut-off  $\omega_s > .70$ ) supporting scale multidimensionality. In the hierarchical model, the general factor explained a large proportion of variance in score (89.0%), with low

**Table 2.** Demographic and clinical characteristics of the total sample as well as exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) sub-samples

Demographic characteristics ( <i>n</i> = 1398)	<i>n</i> (%) / <i>M</i> ( <i>SD</i> ) [min, max]		
	Total sample	EFA sample ( <i>n</i> = 698)	CFA sample ( <i>n</i> = 700)
Age	46.26 (16.38) [18, 94]	46.31 (16.55) [18,94]	46.21 (16.23) [18,89]
Gender			
Female	1150 (82.3%)	590 (84.5%)	560 (80.0%)
Male	235 (16.8%)	103 (14.8%)	132 (18.9%)
Other	13 (0.9%)	5 (0.7%)	8 (1.1%)
Ethnicity			
White	1267 (90.6%)	640 (91.7%)	627 (89.6%)
Asian or Asian British	58 (4.1%)	29 (4.2%)	29 (4.1%)
Mixed/multiple ethnic groups	36 (2.6%)	13 (1.9%)	23 (3.3%)
Black or Black British	22 (1.6%)	10 (1.4%)	12 (1.7%)
Other ethnic groups	15 (1.1%)	6 (0.9%)	9 (1.3%)
Employment			
Employed	737 (52.7%)	369 (52.9%)	368 (52.6%)
Retired	250 (17.9%)	122 (17.5%)	128 (18.3%)
Long-term sick or disabled	164 (11.7%)	80 (11.5%)	84 (12.0%)
Student	93 (6.7%)	50 (7.2%)	43 (6.1%)
Other	47 (3.4%)	24 (3.4%)	23 (3.3%)
Unemployed	36 (2.6%)	19 (2.7%)	17 (2.4%)
Unpaid voluntary work	22 (1.6%)	11 (1.6%)	11 (1.6%)
Not working and not receiving benefits	20 (1.4%)	9 (1.3%)	11 (1.6%)
Home-maker/carer	18 (1.3%)	10 (1.4%)	8 (1.1%)
Actively seeking work	7 (0.5%)	3 (0.4%)	4 (0.6%)
Rather not say	4 (0.3%)	1 (0.1%)	3 (0.4%)
Education			
Undergraduate degree	453 (32.4%)	239 (34.2%)	214 (30.6%)
Postgraduate degree	412 (29.5%)	196 (28.1%)	216 (30.9%)
Secondary school qualification	355 (25.4%)	175 (25.1%)	180 (25.7%)
Trade/technical/vocational training	143 (10.2%)	69 (9.9%)	74 (10.6%)
Less than secondary school	27 (1.9%)	13 (1.9%)	14 (2.0%)
No schooling completed	8 (0.6%)	6 (0.9%)	2 (0.3%)
Long-term conditions			
Other	555 (39.7%)	276 (39.5%)	279 (39.9%)
IBD	401 (28.7%)	192 (27.5%)	209 (29.9%)
Chronic pain	264 (18.9%)	126 (18.1%)	138 (19.7%)
Asthma	229 (16.4%)	113 (16.2%)	116 (16.6%)
Hypertension	179 (12.8%)	84 (12.0%)	95 (13.6%)
Diabetes	173 (12.4%)	92 (13.2%)	81 (11.6%)
Osteoarthritis	127 (9.1%)	70 (10.0%)	56 (8.0%)
Gynecological conditions (e.g. PCOS, endometriosis)	122 (8.7%)	60 (8.6%)	62 (8.9%)
Lupus	100 (7.2%)	48 (6.9%)	52 (7.4%)
Psoriasis	98 (7.0%)	44 (6.3%)	54 (7.7%)
Chronic kidney disease	97 (6.9%)	47 (6.7%)	50 (7.1%)

(Continued)

Table 2. (Continued)

Demographic characteristics (n = 1398)	n (%) / M(SD) [min, max]		
	Total sample	EFA sample (n = 698)	CFA sample (n = 700)
Osteoporosis	96 (6.9%)	49 (7.0%)	47 (6.7%)
Obesity	75 (5.4%)	36 (5.2%)	39 (5.6%)
Rheumatoid arthritis	74 (5.3%)	43 (6.2%)	31 (4.4%)
Multiple sclerosis	63 (4.5%)	36 (5.2%)	27 (3.9%)
Liver disease	59 (4.2%)	32 (4.6%)	27 (3.9%)
Cancer	57 (4.1%)	32 (4.6%)	25 (3.6%)
Long COVID-19	41 (2.9%)	19 (2.7%)	21 (3.0%)
Coronary heart disease (including angina, MI)	32 (2.3%)	8 (1.2%)	24 (3.4%)
Atrial fibrillation	30 (2.1%)	10 (1.4%)	20 (2.9%)
COPD	27 (1.9%)	13 (1.9%)	14 (2.0%)
Epilepsy	21 (1.5%)	10 (1.4%)	11 (1.6%)
Cerebrovascular disease (including stroke and TIA)	17 (1.2%)	9 (1.3%)	8 (1.1%)
Cognitive and learning disabilities	17 (1.2%)	7 (1.0%)	10 (1.4%)
Heart failure	16 (1.1%)	5 (0.7%)	11 (1.6%)
Parkinson's disease	14 (1.0%)	6 (0.9%)	8 (1.1%)
Peripheral arterial disease or peripheral vascular disease	6 (0.4%)	2 (0.3%)	4 (0.6%)
Alcohol dependence	3 (0.2%)	3 (0.4%)	0 (0.0%)
Substance dependence	3 (0.2%)	1 (0.1%)	2 (0.3%)
HIV/AIDS	2 (0.1%)	0 (0.0%)	2 (0.3%)
Sickle-cell anemia	1 (0.1%)	0 (0.0%)	1 (0.1%)
Viral hepatitis (B&C)	1 (0.1%)	0 (0.0%)	1 (0.1%)
LTC comorbidities	838 (59.9%)	408 (58.5%)	430 (61.4%)
Co-morbid psychiatric diagnoses			
Depression	269 (19.2%)	131 (18.8%)	138 (19.7%)
Anxiety	239 (17.1%)	121 (17.3%)	118 (16.9%)
Severe mental illness	36 (2.6%)	15 (2.2%)	21 (3.0%)
Psychosocial factors			
Depression (PHQ-9)	10.30 (5.81) [0, 24]	10.35 (5.97) [0, 24]	10.24 (5.65) [0, 24]
Anxiety (GAD-7)	8.17 (5.53) [0, 21]	8.31 (5.51) [0, 21]	8.03 (5.55) [0, 21]
Psychological Distress (PHQ-ADS)	18.45 (10.57) [0, 45]	18.62 (10.79) [0, 45]	18.28 (10.35) [0, 44]
Condition-specific distress scores			
Diabetes distress (n = 146)	49.10 (21.07) [17, 101]	49.96 (20.93) [17, 101]	48.13 (21.32) [17, 96]
IBD distress (n = 330)	81.82 (40.33) [0, 162]	80.23 (42.46) [0, 161]	83.36 (38.21) [0, 162]

Note: COPD, chronic obstructive pulmonary disease; GAD-7, Generalized Anxiety Disorder Scale; IBD, inflammatory bowel disease; LTC, long-term condition; MI, myocardial infarction; PCOS, polycystic ovary syndrome; PHQ-ADS, Patient Health Questionnaire Anxiety and Depression Scale; PHQ-9, Patient Health Questionnaire-9; TIA, transient ischemic attack.

$\omega$  values for specific factors ( $\omega_{HS} < 0.50$ ). H values were  $< .80$  for specific factors, indicating poorly defined latent variables in the bifactor model. Although the ECV statistics for the general factor were  $> .8$  indicating high proportions of common variance attributable to the general factor, the PUC statistics  $< .8$  suggests the scale should not be interpreted as unidimensional. Therefore, hierarchical multidimensionality was not supported.

The superiority of the CFA versus ESEM lower order model (correlated factors) was supported by generally larger factor correlations in the CFA and small or non-significant cross-loadings in

the ESEM. Moreover, although the ESEM model demonstrated a marginally better fit, it was not justified, as the improvements were small, despite increasing model complexity substantially.

Supplement 12 of the [Supplementary Material](#) presents the final IRD scale.

### Invariance testing

All levels of gender invariance were reached (Table 3), with no significant differences in factor structure (configural invariance),

**Table 3.** Model fit information for estimated SFI models and invariance testing

Model type	$\chi^2$	df	p-value	RMSEA [90% CI]	CFI	TLI	SRMR	AIC	BIC	$ \lambda_{range} $	$\Delta CFI$	$\Delta RMSEA$	$\Delta SRMR$
<i>Final clinical version (14 items)</i>													
Unidimensional	505.67	77	<.001	.089 (.089, .097)	.909	.893	.047	24935.25	25126.40		–	–	–
Lower order (2 factor) CFA	288.29	76	<.001	.063 (.056, .071)	.955	.946	.034	24674.47	24870.16	F1: .70–.86 F2: .62–.80	–	–	–
Bifactor CFA	182.35	63	<.001	.052 (.043, .061)	.975	.963	.025	24548.69	24803.55	G: .63–.80 F1: .16–.40 F2: .12–.63	–	–	–
Lower order (2 factor) ESEM	198.42	64	<.001	.055 (.046, .064)	.972	.960	.024	24585.44	24835.75	F1: .59–.82 F2: .43–1.1			
Bifactor ESEM	121.38	52	<.001	.044 (.034, .054)	.985	.974	.017	24511.33	24816.25	G: .51–.81 F1: .08–.42 F2: .12–.57	–	–	–
<i>Invariance testing: Gender</i>													
Configural	378.79	154	<.001	.065 (.057, .073)	.953	.944	.051	24353.654	24734.98		0.002	0.002	–0.017
Metric/weak	395.26	165	<.001	.064 (.056, .072)	.951	.946	.053	24343.005	24674.40		0.002	–0.002	–0.002
Scalar/strong (full)	404.94	177	<.001	.061 (.053, .069)	.952	.951	.053	24326.668	24603.58		–0.001	–0.005	0.000

Note: G, general factor; F1, factor one (*intrapersonal distress*); F2, factor two (*interpersonal distress*). Measurement invariance levels were said to be reached if (1) CFI did not deteriorate by >.10 in the more restrictive model, (2)  $\Delta RMSEA < .015$ , and (3)  $\Delta SRMR < .030$  in the configural model and <.010 in the scalar model.

factor loadings (metric invariance), or item intercepts (scalar invariance).

### Validity and reliability testing

Additional psychometric assessment was completed using weighted factor scores in the CFA sample (Table 4). Examination of skewness and kurtosis indicated normality (Hair, Anderson, Babin, & Black, 2010). The number of LTC diagnoses had a small, positive, significant correlation with both factors (*intrapersonal*:  $r = 0.230$ ,  $p < 0.001$ , *interpersonal*:  $r = 0.238$ ,  $p < 0.001$ ; inter-item correlations in Supplement 13 of the Supplementary Material).

Assessments were repeated with summed scores (Supplement 14 of the Supplementary Material) to assess if similar properties were found compared with the weighted scores. Minimal differences were found, supporting the use of sum scoring while maintaining strong psychometric properties.

Mean scores by illness groups are represented in Supplement 15 of the Supplementary Material. Regression models of the subscales and most common LTCs in our sample (see Table 2) were run. LTCs positively associated with the *intrapersonal* score were chronic pain ( $\beta = 0.149$ ,  $p < 0.001$ ), rheumatoid arthritis ( $\beta = 0.118$ ,  $p = 0.002$ ), IBD ( $\beta = 0.107$ ,  $p = 0.006$ ), gynecological conditions ( $\beta = 0.080$ ,  $p = 0.035$ ) and diabetes ( $\beta = 0.079$ ,  $p = 0.042$ ); however, hypertension was negatively associated ( $\beta = -0.104$ ,  $p = 0.016$ ). LTCs positively associated with the *interpersonal* score were IBD ( $\beta = 0.144$ ,  $p < 0.001$ ), gynaecological conditions ( $\beta = 0.105$ ,  $p = 0.005$ ), chronic pain ( $\beta = 0.097$ ,  $p = 0.020$ ), psoriasis ( $\beta = 0.089$ ,  $p = 0.018$ ) and rheumatoid arthritis ( $\beta = 0.080$ ,  $p = 0.039$ ).

### Internal consistency

Both the *intrapersonal* and *interpersonal* subscales demonstrated excellent internal consistency, as measured by both Cronbach  $\alpha$  and McDonald's  $\Omega$  statistics (Table 4).

### Convergent validity

Both the *intrapersonal* and *interpersonal* subscales correlated significantly, positively, and weakly with DDS scores. Both subscales demonstrated large, significant, positive correlations with the IBD-DS, PHQ-ADS, and WSAS scores. Both subscales had moderate to strong, significant, positive correlations with subscales of the CBRQ, with strongest correlations between the embarrassment avoidance subscale and the *interpersonal* subscale, and symptom focusing and catastrophizing and the *intrapersonal* subscale. The IRD scale therefore demonstrated good convergent validity with measures of illness-specific distress, generalized distress, functional impairment, and cognitive and behavioral responses to symptoms (Table 4).

Both subscale scores had positive, significant correlations with the slider scale item, indicating higher factor scores were associated with participants describing their LTC(s) as their primary source of distress.

### Test-retest reliability

Both factors demonstrated excellent test-retest reliability (Table 4), with very strong, positive significant correlations between baseline and follow-up scores (*intrapersonal*:  $r = 0.811$ ,  $p < 0.0001$ ; *interpersonal*:  $r = 0.829$ ,  $p < 0.0001$ ).

### ROC analyses

The results of the ROC analyses are reported in Supplement 16 of the Supplementary Material. The true/gold standard class (cases) used were participants who attributed at least 50% of their psychological distress to their LTC(s). This was indicated by 66.28% of our sample (916/1,382 participants, 16 of the 1,398 [1.14%] participants overall had missing IRD scores), indicating moderately high case prevalence. Figure 1 shows IRD factor scores for cases ( $n = 916$ , 66.28 %) and non-cases ( $n = 466$ , 33.72%) at sensitivity and specificity optimizing cut-points for the *intrapersonal* and *interpersonal* factors (14.5 and 11.5). Demographics and disease



**Table 4.** Descriptive statistics, validity testing, and reliability testing of the weighted IRD subscales

	<i>n</i>	Short version	
		Intrapersonal	Interpersonal
Total score, M(SD)	690	12.160 (4.499)	9.706 (4.778)
Skewedness	690	−0.486	−0.182
Kurtosis	690	2.834	2.355
Correlation with LTC count	690	0.230	0.225
<i>Internal consistency</i>			
Baseline alpha	690	0.905	0.887
Follow-up alpha	572	0.914	0.893
Baseline omega	690	0.906	0.888
Follow-up omega	572	0.920	0.898
<i>Convergent/divergent validity</i>			
Depression (PHQ-8)	676	0.657***	0.639***
Anxiety (GAD-7)	668	0.634***	0.589***
Distress (PHQ-ADS)	676	0.695***	0.661***
Impaired functioning (WSAS)	655	0.590***	0.568***
CBRQ fear avoidance	641	0.386***	0.369***
CBRQ catastrophizing	641	0.651***	0.584***
CBRQ symptom focusing	641	0.647***	0.542***
CBRQ embarrassment avoidance	641	0.533***	0.721***
CBRQ rest avoidance	641	0.495***	0.455***
CBRQ all or nothing	641	0.394***	0.360***
CBRQ damage beliefs	641	0.372***	0.308***
Diabetes distress (DDS)	68	0.371**	0.355**
IBD distress (IBD-DS)	167	0.616***	0.585***
IRD slider	542	0.459***	0.377***
Test–retest reliability	572	0.850***	0.853***

Note: \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ .

LTC, long-term condition; PHQ, Patient Health Questionnaire; GAD, Generalized Anxiety Disorder Scale; PHQ-ADS, Patient Health Questionnaire Anxiety Disorder Scale; WSAS, Work and Social Adjustment Scale; CBRQ, cognitive and behavioural responses to symptoms; DDS, Diabetes Distress Scale; IBD-DS, Inflammatory Bowel Disease Distress Scale; IRD, illness related distress.

predictors of caseness are shown in [Supplement 17](#) of the [Supplementary Material](#). The number of LTCs increased odds of caseness; whereas, older age, Asian ethnicity, and diagnoses of asthma and hypertension reduced odds of caseness.

The area under the ROC curve ([Figure 1](#)) values for each IRD scale version and factor were acceptable (lower  $CI_{95}$ ) to excellent (point estimates, upper  $CI_{95}$ ), with 79–88% of participants correctly classified. Sensitivity and specificity values for optimal cut-points showed that 72–87% of cases, and 66–76% non-cases were correctly classified respectively. Based on Positive and Negative Predictive Values, IRD scale cut-points (*intrapersonal* factor: 14.5, *interpersonal* factor: 11.5) resulted in a correct classification of 82–87% of participants who attributed at least 50% of their psychological distress to their LTC(s) correctly, and of 58–73% of participants who attribute a lower percentage of their distress to their LTC(s). Positive and Negative Likelihood Ratios indicate that cases are

2.11–3.62 times more likely to score above optimal IRD cut-points than non-cases, and non-cases are 2.32–5.88 times more likely to score below optimal IRD cut-points than cases. For the *intrapersonal* factor, the optimal cut-point of 14.5 identified 97.82% cases (896/916 participants) correctly, and 27.90% non-cases incorrectly (130/466 participants). For the *interpersonal* factor, the optimal cut-point of 11.5 identified 92.03% cases (843/916 participants), and 29.97% non-cases incorrectly (135/466 participants).

### Readability

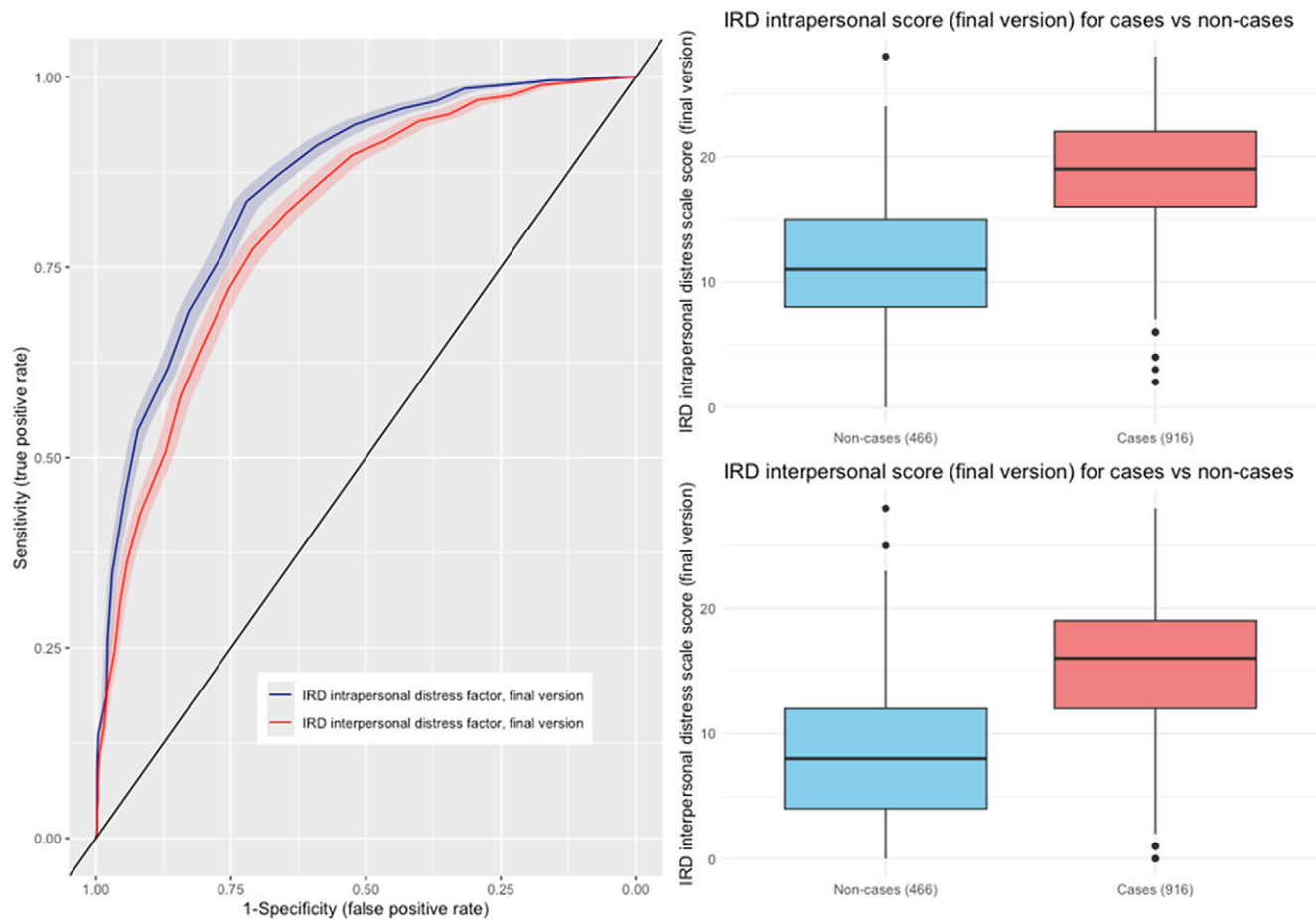
The final IRD scale ([Supplement 12](#) of the [Supplementary Material](#)) had a Flesch Reading Ease score of 74.9, equivalent to US grade 5 reading level of 10–11 years old.

### Discussion

This study aimed to develop a measure of Illness-Related Distress that can be used across LTCs and to test the psychometric properties of this new scale. To our knowledge, this is the first transdiagnostic measure of IRD. The final IRD Scale was comprised of two 7-item factors demonstrated through EFA and confirmed by CFA. Model fitting demonstrated that a single factor or bifactor model was not supported. A CFA lower-order correlated factor model was favored over ESEM due to marginal differences in fit statistics, and greater simplicity with the CFA. Therefore, the two factors, although conceptually related, should be calculated separately and not combined into a total score. The model had excellent fit statistics and passed invariance testing.

We labeled the two factors the *intrapersonal distress* subscale, measuring a range of emotions directly related to the challenges of living with an LTC such as anger, frustration, and worry, and the *interpersonal distress* subscale, capturing feelings associated with social/self-perception issues, such as being embarrassed by the illness or feeling like a burden. The subscales demonstrated excellent internal reliability, good test–retest reliability, good readability, and promising clinical cut points against the reference category (asking the percentage of distress respondents attributed to their LTC(s)). There were significant, small to large positive correlations between the subscales and measures of conceptually related constructs, including psychological distress, depression, anxiety, impaired functioning, cognitive and behavioral responses to symptoms, and illness-specific measures of distress (in diabetes and IBD). While convergent validity was supported with moderately strong correlations between the IRD subscales and IBD-related distress, the relationships between IRD subscales and diabetes-related distress were smaller. This may be because the Diabetes Distress Scale (DDS) includes items concerning impact and management (e.g. ‘Feeling that my doctor doesn’t give me clear enough directions on how to manage by diabetes’), rather than purely distress. Overall, the results indicate that the IRD scale is a brief, valid, reliable, and potentially clinically informative instrument for measuring and classifying transdiagnostic IRD.

Extensive research underpinned the initial scale item pool, including qualitative interviews with people living with LTCs, expert consensus meetings, a systematic literature search, and feedback on items from people living with LTCs and HCPs who treat anxiety and depression in LTCs. This focus on the face validity of items may underly the robust psychometrics of the scale (Boateng, Neilands, Frongillo, Melgar-Quinonez, & Young, 2018;



**Figure 1.** Receiver operating characteristic (ROC) curves and associated 95% confidence intervals (left panel) for the illness related distress (IRD) scale, final version (*intrapersonal* and *interpersonal* factors) as a predictor of attributing  $\geq 50\%$  of psychological distress to the primary long-term condition. Boxplots of underlying IRD scores by caseness are shown in the right panel. Note: IRD, 'illness related distress'.

Flake & Fried, 2020; Morgado, Meireles, Neves, Amaral, & Ferreira, 2017). Given its excellent psychometrics and good readability, the IRD Scale may be an acceptable tool to introduce consistency in the measurement of distress related to adjusting to LTCs, irrespective of diagnosis and multimorbidity.

There are several potential clinical applications of this scale. First, it can improve understanding of how IRD is experienced, incorporating a wider range of emotions beyond depression and anxiety. Second, it may help clinicians determine the presenting problem, that is, whether current distress is related to the difficulties of adjusting to the challenges of an LTC (IRD), or if the presentation resembles primary anxiety or depression. The range for each IRD subscale is 0–28. Preliminary analysis suggests cut-offs of 14.5 and 11.5, respectively, for the *intrapersonal* and *interpersonal* factors of the scale, rounded to 15 and 12 for clinical use. Importantly, these cut-offs do not indicate whether distress is clinically significant or preclude a prior or primary diagnosis that may increase vulnerability to IRD. However, they can be used to decide if a significant proportion of distress is LTC-related, thus signaling whether a therapy tailored to IRD may be most appropriate. Third, it may help to identify specific treatment targets for interventions. For example, those who score high on interpersonal distress may benefit from interventions that focus on making social connections to reduce loneliness, challenging cognitions related to embarrassment, and exploring ways to reciprocate social support when feeling

like a burden. Finally, the IRD Scale could be used as a primary outcome measure in trials assessing interventions designed to treat IRD and adjustment in LTCs. This could be utilized alongside more traditional measures of anxiety and depression such as the GAD-7 or PHQ-9, to explore relative sensitivity to change.

### Strengths, limitations, and future directions

The IRD scale was developed rigorously, ensuring common pitfalls in scale development were avoided by (1) specifying a construct; (2) confirming the absence of existing scales through a literature search; (3) prioritizing lived experience by conducting exploratory interviews with and getting detailed feedback from LTC patients; and (4) consulting expert judges (clinicians) (Boateng *et al.*, 2018). The study used a large sample with diverse LTCs and employed sophisticated analyses to reduce item pool and identify best model fit, allowing the assessment of varied complex structures and construct-relevant multidimensionality. Moreover, scale development relied upon factor loadings, factor correlations, and general factor structure, alongside health psychology theory.

Despite the large sample size, the self-selected community sample may limit generalizability. Overrepresented demographics included white ethnicity, female gender, and high educational level. Moreover, some LTCs appeared to be underrepresented based on

epidemiological prevalence (e.g. hypertension, obesity, type 2 diabetes). However, the subscales passed invariance testing (based on gender), suggesting that this demographic does not impact factor structure. Although internet research allows anonymous participation, improved accessibility, minimized embarrassment, social stigma, and fear of judgment, it prevents confirmation of LTC diagnoses. The ROC analysis, used to define the clinical cut-points has limitations, as the definition used for a case was a single item non-validated ratio rather than a severity measure (Pepe, Janes, Longton, Leisenring, & Newcomb, 2004). However, as there is no gold standard measure of IRD, the ratio measure provides a starting point. Future research should systematically compare methods and explore machine learning approaches for potentially improved accuracy. Moreover, the model could be tested for invariance based on LTC diagnosis and/or with qualitatively different EFA and CFA/ESEM samples instead of different data portions. Though the IRD scale measures the severity of distress, a transdiagnostic illness-related stressor checklist may also complement research and clinical decision-making in this area. Although initial convergent validity was performed in this study, future research should include measures to test divergent validity. Finally, sensitivity to change and criterion validity should be assessed by utilizing the IRD subscales in intervention studies and comparing IRD cut-offs with diagnostic interviews.

## Conclusions

The IRD scale is a 14-item valid and reliable measure, comprised of two factors of distress (*intrapersonal* and *interpersonal* IRD). It reliably captures IRD, with excellent evidence of internal consistency, convergent validity with thematically similar measures, and test–retest reliability. The IRD scale has significant clinical utility in clinical and research settings, particularly in treatment decision-making and the assessment of treatment efficacy. Further research is needed to assess sensitivity to change and criterion validity.

**Supplementary material.** The supplementary material for this article can be found at <http://doi.org/10.1017/S003329172500090X>.

**Acknowledgements.** We would like to thank the following charities who advertised our study with their network: Alopecia UK, Arthritis Action UK, The Arthritis and Musculoskeletal Alliance (ARMA), Asthma + Lung UK, Bowel Research UK, Brave Hearts Northern Ireland, Breast Cancer Now, Breathe Arts Health Research, British Polio Fellowship, British Porphyria Association, British Skin Foundation, Burning Nights, Crohn's & Colitis UK, Complex Regional Pain Syndrome (CRPS) UK, Dewi Cymru, Diabetes UK, Fibromyalgia Action UK, Juvenile Diabetes Research Foundation, Kidney Care UK, Kidney Research UK, Lipoedema UK, Liver4lifeUK, Motor Neuron Disease Association, Multiple Sclerosis (MS) Trust, National Kidney Federation, National Rheumatoid Arthritis Society (NRAS) UK, Neuroblastoma UK, Neuroendocrine Cancer UK, Pain UK, The Psoriasis and Psoriatic Arthritis Alliance, Psoriasis UK, Royal Osteoporosis Society, Stroke Association. We would like to thank students Edward Le Marchant, Amber (Kai Xin) Siow, and Glenn (Chin) Kong for their support with study recruitment.

**Funding statement.** This study has been delivered through the National Institute for Health and Care Research (NIHR) Maudsley Biomedical Research Centre (BRC) (NIHR203318). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

**Competing interests.** R.M.M. has received personal fees from Mahana Therapeutics for scientific advisory work and from other universities and hospital trusts for cognitive behavioral therapy training in irritable bowel syndrome. She was a beneficiary of a license agreement between Mahana

Therapeutics and King's College London. The remaining authors declare no further competing interests.

## References

- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, **52**, 317–332.
- Allen, M. S., Iliescu, D., & Greiff, S. (2022). *Single item measures in psychological science*. Hogrefe Publishing.
- Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, **16**(3), 397–438.
- Ayers, S., & Steptoe, A. (2007). Stress and health. In S. Ayers, A. Baum, C. McManus, S. Newman, K. Wallston, J. Weinman, & R. West (Eds.), *Cambridge handbook of psychology, health and medicine* (2nd ed., pp. 215–219). Cambridge: Cambridge University Press.
- Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *BMJ*, **314**(7080), 572.
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, **6**, 149.
- Brown, A., Barker, E. D., & Rahman, Q. (2022). Development and psychometric validation of the sexual fantasies and behaviors inventory. *Psychological Assessment*, **34**(3), 217.
- Browne, J. L., Ventura, A., Mosely, K., & Speight, J. (2013). 'I call it the blame and shame disease': A qualitative study about perceptions of social stigma surrounding type 2 diabetes. *BMJ Open*, **3**(11), e003384. <https://doi.org/10.1136/bmjopen-2013-003384>
- Carroll, S., Moon, Z., Hudson, J., Hulme, K., & Moss-Morris, R. (2022). An evidence-based theory of psychological adjustment to long-term physical health conditions: Applications in clinical practice. *Psychosomatic Medicine*, **84**(5), 547–559. <https://doi.org/10.1097/psy.0000000000001076>
- Carroll, S., Moss-Morris, R., Hulme, K., & Hudson, J. (2021). Therapists' perceptions of barriers and facilitators to uptake and engagement with therapy in long-term conditions. *British Journal of Health Psychology*, **26**(2), 307–324.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, **14**(3), 464–504. <https://doi.org/10.1080/10705510701301834>
- Cliff, N. (1988). The eigenvalues-greater-than-one rule and the reliability of components. *Psychological Bulletin*, **103**(2), 276.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic Press.
- Costello, A. B., & Osborne, J. (2019). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, **10**(1), 7.
- Coulter, A., Entwistle, V. A., Eccles, A., Ryan, S., Shepperd, S., & Perera, R. (2015). Personalised care planning for adults with chronic or long-term health conditions. *Cochrane Database of Systematic Reviews*, **2015**(3), CD0105. <https://doi.org/10.1002/14651858>
- Cuvillier, M., Léger, P.-M., & Sénécal, S. (2021). Quantity over quality: Do single-item scales reflect what users truly experienced? *Computers in Human Behavior Reports*, **4**, 100097.
- Diabetes UK. (2019). *Too often missing: Making emotional and psychological routine in diabetes care*. London: Diabetes UK. <https://www.diabetes.org.uk/support-us/campaign/other-campaigns/its-missing-evidence>
- Dibley, L., Czuber-Dochan, W., Woodward, S., Wade, T., Bassett, P., Sturt, J., & Norton, C. (2018). Development and psychometric properties of the inflammatory bowel disease distress scale (IBD-DS): A new tool to measure disease-specific distress. *Inflammatory Bowel Diseases*, **24**(9), 2068–2077. <https://doi.org/10.1093/ibd/izy108>
- Dicke, T., Marsh, H. W., Riley, P., Parker, P. D., Guo, J., & Horwood, M. (2018). Validating the copenhagen psychosocial questionnaire (COPSOQ-II) using set-ESEM: Identifying psychosocial risk factors in a sample of school principals. *Frontiers in Psychology*, **9**, 584.
- Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency



- estimation. *British Journal of Psychology*, **105**(3), 399–412. <https://doi.org/10.1111/bjop.12046>
- Ellison, L., Gask, L., Bakerly, N. D., & Roberts, J. (2012). Meeting the mental health needs of people with chronic obstructive pulmonary disease: A qualitative study. *Chronic Illness*, **8**(4), 308–320.
- Fisher, L., Glasgow, R. E., Mullan, J. T., Skaff, M. M., & Polonsky, W. H. (2008). Development of a brief diabetes distress screening instrument. *Annals of Family Medicine*, **6**(3), 246–252. <https://doi.org/10.1370/afm.842>
- Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, **3**(4), 456–465.
- Fleetwood, K. J., Guthrie, B., Jackson, C. A., Kelly, P. A., Mercer, S. W., Morales, D. R., ... Prigge, R. (2025). Depression and physical multimorbidity: A cohort study of physical health condition accrual in UK Biobank. *PLoS Medicine*, **22**(2), e1004532.
- Geraghty, K. J., & Esmail, A. (2016). Chronic fatigue syndrome: Is the biopsychosocial model responsible for patient dissatisfaction and harm? *British Journal of General Practice*, **66**(649), 437–438. <https://doi.org/10.3399/bjgp16X686473>
- Hair, J. F., Anderson, R. E., Babin, B. J., & Black, W. C. (2010). *Multivariate data analysis: A global perspective* (Vol. 7). Upper Saddle River, NJ: Pearson.
- Head, A., Fleming, K., Kypridemos, C., Schofield, P., Pearson-Stuttard, J., & O'Flaherty, M. (2021). Inequalities in incident and prevalent multimorbidity in England, 2004–19: A population-based, descriptive study. *The Lancet Healthy Longevity*, **2**(8), e489–e497.
- Hooper, H., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods*, **6**(1), 53–60. <https://academic-publishing.org/index.php/ejbrm/article/view/1224>
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, **30**, 179–185.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, **6**(1), 1–55.
- Hudson, J. L., & Moss-Morris, R. (2019). Treating illness distress in chronic illness: Integrating mental health approaches with illness self-management. *European Psychologist*, **24**(1), 26–37. <https://doi.org/10.1027/1016-9040/a000352>
- IBD UK. (2021). *Crohn's and colitis care in the UK: the hidden cost and a vision for change*. IBD UK.
- Jenkinson, E., Hudson, J., Moss-Morris, R., & Hackett, R. (in prep.). Stakeholder perspectives on implementing internet-based Cognitive Behavioural Therapy (CBT) into routine clinical practice for adults with diabetes and psychological distress.
- Katon, W., & Roy-Byrne, P. P. (1991). Mixed anxiety and depression. *Journal of Abnormal Psychology*, **100**(3), 337–345. <https://doi.org/10.1037/0021-843X.100.3.337>
- Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S.-L., ... Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, **32**(6), 959–976.
- Knies, G., & Kumari, M. (2022). Multimorbidity is associated with the income, education, employment and health domains of area-level deprivation in adult residents in the UK. *Scientific Reports*, **12**(1), 7280. <https://doi.org/10.1038/s41598-022-11310-9>
- Kreider, K. E. (2017). Diabetes distress or major depressive disorder? A practical approach to diagnosing and treating psychological comorbidities of diabetes. *Diabetes Therapy*, **8**(1), 1–7. <https://doi.org/10.1007/s13300-017-0231-1>
- Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9 – Validity of a brief depression severity measure. *Journal of General Internal Medicine*, **16**(9), 606–613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>
- Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B., Berry, J. T., & Mokdad, A. H. (2009). The PHQ-8 as a measure of current depression in the general population. *Journal of Affective Disorders*, **114**(1–3), 163–173.
- Kroenke, K., Wu, J., Yu, Z., Bair, M. J., Kean, J., Stump, T., & Monahan, P. O. (2016). The patient health questionnaire anxiety and depression scale (PHQ-ADS): Initial validation in three clinical trials. *Psychosomatic Medicine*, **78**(6), 716.
- Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit in structural equation. *Contemporary Psychometrics*.
- Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, **10**, 85–110. <https://doi.org/10.1146/annurev-clinpsy-032813-153700>
- McDaid, D., Knapp, M., Fossey, M., & Galea, A. (2012). Long-term conditions and mental health. In: *Centre For Mental Health*, The King's Fund.
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., ... de Vet, H. C. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, **63**(7), 737–745.
- Morgado, F. F., Meireles, J. F., Neves, C. M., Amaral, A. C., & Ferreira, M. E. (2017). Scale development: Ten main limitations and recommendations to improve future research practices. *Psicologia: Reflexão e Crítica*, **30**(0), 3.
- Morin, A. J. S., Myers, N. D., & Lee, S. (2020). Modern factor analytic techniques. In Bifactor models, exploratory structural equation modeling (ESEM), and bifactor-ESEM. In *Handbook of sport psychology*, (eds G. Tenenbaum and R.C. Eklund). <https://doi.org/10.1002/9781119568124.ch51> (pp. 1044–1073).
- Morin, A. J. S., Arens, A. K., Tran, A., & Caci, H. (2016). Exploring sources of construct-relevant multidimensionality in psychiatric measurement: A tutorial and illustration using the composite scale of morningness. *International Journal of Methods in Psychiatric Research*, **25**(4), 277–288. <https://doi.org/10.1002/mpr.1485>
- Mossadeghi, B., Caixeta, R., Ondarsuhu, D., Luciani, S., Hambleton, I. R., & Hennis, A. J. M. (2023). Multimorbidity and social determinants of health in the US prior to the COVID-19 pandemic and implications for health outcomes: a cross-sectional analysis based on NHANES 2017–2018. *BMC Public Health*, **23**(1), 887. <https://doi.org/10.1186/s12889-023-15768-8>
- Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. M. (2002). The work and social adjustment scale: A simple measure of impairment in functioning. *The British Journal of Psychiatry*, **180**(5), 461–464.
- NICE. (2024). *Browse guidance by topic: Conditions and diseases*. NICE. <https://www.nice.org.uk/guidance/conditions-and-diseases>
- Nunnally, J. C., & Bernstein, I. (1978). *Psychometric theory* mcgraw-hill new york. The role of university in the development of entrepreneurial vocations: A Spanish study, 387–405.
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, **32**(3), 396–402.
- Office for National Statistics. (2022). *UK health indicators: 2019 to 2020*. ONS. <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/bulletins/ukhealthindicators/2019to2020>
- Pepe, M. S., Janes, H., Longton, G., Leisenring, W., & Newcomb, P. (2004). Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *American Journal of Epidemiology*, **159**(9), 882–890.
- Picariello, F., Chilcot, J., Chalder, T., Herdman, D., & Moss-Morris, R. (2023). The Cognitive and Behavioural Responses to Symptoms Questionnaire (CBRQ): Development, reliability and validity across several long-term conditions. *British journal of health psychology*, **28**(2), 619–638. <https://doi.org/10.1111/bjhp.12644>
- Ponzo, M., Tacchino, A., Zaratini, P., Vaccaro, C., & Battaglia, M. A. (2015). Unmet care needs of people with a neurological chronic disease: A cross-sectional study in Italy on multiple sclerosis. *The European Journal of Public Health*, **25**(5), 775–780.
- Read, J. R., Sharpe, L., Modini, M., & Dear, B. F. (2017). Multimorbidity and depression: A systematic review and meta-analysis. *Journal of Affective Disorders*, **221**, 36–46.



- Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354.
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*, 1–8.
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137–150. <https://doi.org/10.1037/met0000045>
- Salazar, M. S. (2015). The dilemma of combining positive and negative items in scales. *Psicothema*, *27*(2), 192–199.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.
- Schwarz, T., Schmidt, A. E., Bobek, J., & Ladurner, J. (2022). Barriers to accessing health care for people with chronic conditions: A qualitative interview study. *BMC Health Services Research*, *22*(1), 1037.
- Spitzer, R., Kroenke, K., & Williams, J. (2006). Generalized anxiety disorder 7-item (GAD-7) scale. *Archives of Internal Medicine*, *166*, 1092–1097.
- Stewart-Knight, K., Parry, R., Abey, A., & Seymour, J. (2012). Does the distress thermometer work? A systematic review of the evidence for its use and validity. *BMJ Supportive & Palliative Care*, *2*(Suppl 1), A30–A30.
- Watkins, M. W. (2013). *Omega*. Pheonix: Ed & Psych Associates.
- Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, *51*(1), 409–428. <https://doi.org/10.3758/s13428-018-1055-2>
- Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*(1), 32–35.