Manuscript: 5,092 words
Abstract: 244 / 250 words
Figures: 2
Tables: 3
References: 64
eTables: 7
eFigures: 4

Validation of the Collaborative Outcomes study on Health and Functioning during Infection Times (COH-FIT) questionnaire for adults

Marco Solmi[1,2,3,4,5,6]*, Trevor Thompson[7]*, Andrés Estradé[4,8], Agorastos Agorastos[9], Joaquim Radua[4,10,11], Samuele Cortese[12], Elena Dragioti[13,14], Friedrich Leisch[15], Davy Vancampfort[16], Lau Caspar Thygesen[17], Harald Aschauer[18], Monika Schloegelhofer[18], Elena Aschauer[18], Andres Schneeberger[19], Christian G. Huber[20], Gregor Hasler[21], Philippe Conus[22], Kim Q. Do Cuénod[22], Roland von Känel[23], Gonzalo Arrondo[12,24], Paolo Fusar-Poli[4,25,26], Philip Gorwood[27,28], Pierre-Michel Llorca[29], Marie-Odile Krebs[28,30], Elisabetta Scanferla[27], Taishiro Kishimoto[31], Golam Rabbani[32], Karolina Skonieczna-Żydecka[33], Paolo Brambilla[34,35], Angela Favaro[36], Akihiro Takamiya[31], Leonardo Zoccante[37], Marco Colizzi[38], Julie Bourgin[39], Karol Kamiński[40], Maryam Moghadasin[41], Soraya Seedat[42], Evan Matthews[43], John Wells[43], Emilia Vassilopoulou[44], Ary Gadelha[45], Kuan-Pin Su[46,114], Jun Soo Kwon[47], Minah Kim[48], Tae Young Lee[49], Oleg Papsuev[50], Denisa Manková[51], Andrea Boscutti[34], Cristiano Gerunda[36], Diego Saccon[52], Elena Righi[53], Francesco Monaco[54], Giovanni Croatto[55], Guido Cereda[34], Jacopo Demurtas[56], Natascia Brondino[26], Nicola Veronese[57], Paolo Enrico[34], Pierluigi Politi[26], Valentina Ciappolino[35], Andrea Pfennig[58], Andreas Bechdolf[59], Andreas Meyer-Lindenberg[60], Kai G. Kahl[61], Katharina Domschke[62], Michael Bauer[58], Nikolaos Koutsouleris[63], Sibylle Winter[6], Stefan Borgwardt[64], Istvan Bitter[65], Judit Balazs[66,67], Pal Czobor[65], Zsolt Unoka[65], Dimitris Mavridis[68], Konstantinos Tsamakis[69], Vasilios P. Bozikas[9], Chavit Tunvirachaisakul[70], Michael Maes[70], Teerayuth Rungnirundorn[70], Thitiporn Supasitthumrong[70], Ariful Haque[32], Andre R. Brunoni[71], Carlos Gustavo Costardi[45], Felipe Barreto Schuch[72], Guilherme Polanczyk[71], Jhoanne Merlyn Luiz[73], Lais Fonseca[45], Luana V. Aparicio[71], Samira S. Valvassori[73], Merete Nordentoft[74], Per Vendsborg[75], Sofie Have Hoffmann[17], Jihed Sehli[21], Norman Sartorius[76], Sabina Heuss[77], Daniel Guinart[78,79,80], Jane Hamilton[81], John Kane[78,82], Jose Rubio[78,82], Michael Sand[83], Ai Koyanagi[84], Aleix Solanes[10], Alvaro Andreu-Bernabeu[85], Antonia San José Cáceres[85], Celso Arango[85], Covadonga M. Díaz-Caneja[85], Diego Hidalgo-Mazzei[86], Eduard Vieta[86], Javier Gonzalez-Peñas[85], Lydia Fortea[10], Mara Parellada[85], Miquel A. Fullana[10], Norma Verdolini[87], Eva Andrlíková [51], Karolina Janků[51], Mark John Millan[88], Mihaela Honciuc[29], Anna Moniuszko-Malinowska[89], Igor Łoniewski[33,90], Jerzy Samochowiec[91], Łukasz Kiszkiel[92], Maria Marlicz[33], Paweł Sowa[40], Wojciech Marlicz[93,94], Georgina Spies[42], Brendon Stubbs[95], Joseph Firth[96], Sarah Sullivan[97], Asli Enez Darcin[98], Hatice Aksu[99], Nesrin Dilbaz[100], Onur Noyan[100], Momoko Kitazawa[31], Shunya Kurokawa[31], Yuki Tazawa[31], Alejandro Anselmi[8], Cecilia Cracco[8], Ana Inés Machado[8], Natalia Estrade[8], Diego De Leo[101], Jackie Curtis[102], Michael Berk[103], Philip Ward[104], Scott Teasdale[103], Simon Rosenbaum[104], Wolfgang Marx[103], Adrian Vasile Horodnic[105], Liviu Oprea[105], Ovidiu Alexinschi[106], Petru Ifteni[107], Serban Turliuc[105], Tudor Ciuhodaru[108], Alexandra Bolos[105], Valentin Matei[109], Dorien H. Nieman[110], Iris Sommer[111,112], Jim van Os[113], Therese van Amelsvoort[114], Ching-Fang Sun[115,116], Ta-wei Guu[117], Can Jiao[118], Jieting Zhang[118], Jialin Fan[118], Liye Zou[118], Xin Yu[119], Xinli Chi[118], Philippe de Timary[120], Ruud van Winkel[121], Bernardo Ng[122], Edilberto Pena[122], Ramon Arellano[122], Raquel Roman[122], Thelma Sanchez[122], Larisa Movina[50], Pedro Morgado[123,124], Sofia Brissos[125], Oleg Aizberg[126], Anna Mosina[127], Damir Krinitski[128], James Mugisha[129], Dena Sadeghi-Bahmani[130,131], Farshad Sheybani[132], Masoud Sadeghi[133], Samira Hadi[134], Serge Brand[131,135,136,137,138], Antonia Errazuriz[139], Nicolas Crossley[139], Dragana Ignjatovic Ristic[140], Carlos López-Jaramillo[141], Dimitris Efthymiou[44], Praveenlal Kuttichira[142], Roy Abraham Kallivayalil[143], Afzal Javed[144], Muhammad Iqbal Afridi[145,146,147], Bawo James[148], Omonefe Joy Seb-Akahomen[149], Jess Fiedorowicz[1,2,3], Andre F. Carvalho[103], Jeff Daskalakis[150], Lakshmi N. Yatham[151], Lin Yang[152], Tarek Okasha[153], Aïcha Dahdouh[154], Björn Gerdle[13], Jari Tiihonen[11], Jae Il Shin[155], Jinhee Lee[156], Ahmed Mhalla[157], Lotfi Gaha[157], Takoua Brahim[158], Kuanysh Altynbekov[159], Nikolay Negay[159], Saltanat Nurmagambetova[159], Yasser Abu Jamei[160], Mark Weiser[161], Christoph U. Correll[6,78,79][§]

* Joint first authors

[§] **Corresponding author:** Christoph U. Correll, MD, Department of Child and Adolescent Psychiatry, Psychosomatic Medicine and Psychotherapy, Charité University Medical Center, Campus Virchow, Augustenburger Platz 1, D-13353, Berlin, Germany. Tel.: +49 30 450 566202 , Fax: +49 30 450 566921 , E-mail: christoph.correll@charite.de

Abstract 249/250

Background. The Collaborative Outcome study on Health and Functioning during Infection Times (COH-FIT; www.coh-fit.com) is an anonymous and global online survey measuring health and functioning during COVID-19 pandemic. The aim of this study was to test concurrently the validity of COH-FIT items and the internal validity of the co-primary outcome, a composite psychopathology "P-score".

Methods. The COH-FIT survey has been translated into 30 languages (two blind forward-translations, consensus, one independent English back-translation, final harmonization). To measure mental health, 1-4 items ("COH-FIT items") were extracted from validated questionnaires (e.g. Patient Health Questionnaire 9). COH-FIT items measured anxiety, depressive, post-traumatic, obsessive-compulsive, bipolar and psychotic symptoms, as well as stress, sleep and concentration. COH-FIT Items which correlated $r \geq 0.5$ with validated companion questionnaires, were initially retained. A P-score factor structure was then identified from these items using exploratory factor analysis (EFA) and confirmatory factor analyses (CFA) on data split into training and validation sets. Consistency of results across languages, gender and age was assessed.

Results. From >150,000 adult responses by May 6th, 2022, a subset of 22,456 completed both COH-FIT items and validated questionnaires. Concurrent validity was consistently demonstrated across different languages for COH-FIT items. CFA confirmed EFA results of five first-order factors (anxiety, depression, post-traumatic, psychotic, psychophysiologic symptoms) and revealed a single second-order factor P-score, with high internal reliability ($\omega=0.95$). Factor structure was consistent across age and sex.

Conclusions. COH-FIT is a valid instrument to globally measure mental health during infection times. The P-score is a valid measure of multidimensional mental health.

Keywords

Covid-19; pandemic; COH-FIT; survey: P-factor: well-being: mental health: psychiatry: psychometric

Introduction

COVID-19 has infected over 530 million people and caused almost 6.3 million deaths up to June 1st, 2022, since its breakout, globally(Dong et al., 2020). The indirect impact of COVID-19 on mental health of the general population(Dragioti et al., 2021) and of specific groups(Chen et al., 2022; Dragioti et al., 2022; Leung et al., 2022; Zhang et al., 2022) of the population has been studied by several anonymous surveys. Mental health surveys published in the early stage of the pandemic recruited on average 5,137 respondents and a maximum of 56,679 respondents(Lin et al., 2021), in adults. In children and adolescents(Theberath et al., 2022), surveys focused mainly on anxiety (28%) and depression (23%), while loneliness (5%), stress (5%), fear (5%), tension (3%), anger (3%), fatigue (3%), confusion (3%), and worry (3%) were assessed much less frequently. Most surveys focused on a few outcomes. The largest meta-analysis on the prevalence of mental health outcomes during the COVID-19 pandemic, which included 173 surveys and over 500,000 participants, showed that the highest prevalence during the COVID-19 pandemic is for posttraumatic symptoms in COVID-19-infected people (94%), but also that mental health can be broadly affected by the COVID-19 pandemic. These outcomes included behavioral problems in those with prior mental disorders (77%), fear in healthcare workers (71%), anxiety in caregivers/relatives of people infected with COVID-19 (42%), general health/social contact/passive coping style in the general population (38%), depression in those with prior somatic disorders (37%), and fear in other-than-healthcare workers (29%)(Dragioti et al., 2021). Females seem to be particularly affected by the pandemic overall, college students/young adults with respect to anxiety, depressive and sleep problems, and suicidal ideation, and adults with regards to post-traumatic stress disorder(Dragioti et al., 2021).

Given the evidence of the multidimensional impact of the pandemic on mental health in the general population, surveys ideally should assess a composite psychopathology domain: “p”, which covers and incorporates these different aspects of mental health and functioning. Numerous studies have shown that the many psychiatric symptoms and disorders ultimately cluster in three psychopathology dimensions (namely, externalizing, internalizing, and psychotic experiences), which in turn load on a single domain of psychopathology, “p”, paralleling the “g” factor for intelligence, and mapping on a continuum from low to extreme psychopathology(Caspi et al., 2014; Caspi and Moffitt, 2018). Importantly, externalizing behaviour is difficult to capture and measure with online surveys, while valid self-report questionnaires exist for internalizing and psychotic symptoms. P is classically conceived as a latent variable, putatively associated with an increased risk of developing mental disorders(Caspi et al., 2014; Caspi and Moffitt, 2018). However, it has been proposed that P should also be considered as a mental health outcome in clinical studies aiming to prevent or treat mental disorders(Caspi and Moffitt, 2018). Accounting for the dual nature of P, as a latent vulnerability factor and as an outcome, a questionnaire measuring composite psychopathology could inform on both vulnerability for future development of mental disorders (P as a liability latent factor), and the

broad mental health status (P as an outcome). To assess P, measures of individual psychopathological domains are needed. Most of the surveys conducted during the COVID-19 pandemic to date have focused on one or two psychopathology domains, and have used full-length validated questionnaires, that are composed of numerous items. This approach has limited the number of domains that could be covered within a reasonable amount of time. For instance, among others, the Patient Health Questionnaire 9 (PHQ-9)(Kroenke et al., 2001) was frequently used to measure depressive symptoms, the Generalized Anxiety Disorder 7 (GAD-7)(Spitzer et al., 2006) to measure anxiety symptoms, and the post-traumatic stress disorder (PTSD) Checklist for DSM-5 (PCL-5)(Blevins et al., 2015) to measure post-traumatic symptoms. These questionnaires are nine, seven, and 20 items long, respectively. Hence, in the context of an online survey, using validated questionnaires to assess P and create a P-score would take too many items, likely decreasing completion rates of responses. An alternative approach to creating a P-score is to use fewer items to measure multiple dimensions of mental health at the same time, minimising time demands and avoiding fatigue of the participant.

It is also very important to keep in mind that online surveys are not limited by borders, and that they have the potential of reaching people living in any country and speaking any language. However, almost every online survey normally provides the option to answer in one or (rarely) two languages, most frequently English, or Chinese. This limitation is of particular concern as it can introduce selection bias since the pandemic is particularly affecting the most fragile strata of the population, including ethnic and linguistic minorities who generally have lower socio-economic status and education(Pan et al., 2020; Treweek et al., 2020), and who are frequently non-fluent in the official national language of the country of residence(UK_Government, 2011). Hence, a multi-language survey has the potential of being more inclusive, not leaving behind any linguistic minorities, and collect evidence from as many countries globally as possible(Solmi et al., 2021).

However, the use of abbreviated scales to measure mental health requires evidence that the scale validity is not adversely affected. Furthermore, merging item data from the same survey across multiple language translations assumes that the psychometric properties are not compromised by their presentation in a different language.

The Collaborative Outcome study on Health and Functioning during Infection Times (COH-FIT; www.coh-fit.com) is an online survey measuring the impact of COVID-19 pandemic on health and functioning of the general populations. COH-FIT is one of the largest international, multi-language (n=30), cross-sectional, anonymous online surveys for adults, adolescents (14-17 years), and children (6-13 years), measuring health and functioning during COVID-19 pandemic globally in a multi-wave design, utilizing both non-probability and representative sampling, in collaboration with over 220 researchers from all around the globe(Solmi et al., 2022b, 2022c). Since April 26th, 2020 up to May 6th, 2022, COH-FIT has collected over 150,000 responses from adults and over 15,000 responses from minors, in over 150 countries. The design of COH-FIT has been described and discussed in detail

previously(Solmi et al., 2022b, 2022c, 2022a). Briefly, COH-FIT assesses at the time of taking the survey - and recalled for the last 2 weeks of regular life before the onset of the pandemic locally - aspects of both physical health and mental health in order to measure the impact of the pandemic, including its relationship to specific moderators and mediators of that impact. With regards to the assessment of mental health, COH-FIT, uses selected items for each psychopathology domain that were extracted from full-length validated questionnaires, which are then put together to build a composite general psychopathology P-score. The COH-FIT P-score is composed by COH-FIT items that are found to sufficiently represent that full validated scale result for anxiety, depressive, post-traumatic, obsessive-compulsive, bipolar and psychotic symptoms, as well as psychophysiologic measures of stress, sleep, and concentration problems. The primary aim of this validation study was to evaluate the psychometric properties of the COH-FIT P-score by (1) examining the concurrent validity of each of the selected COH-FIT psychopathology items and domains, via examining correlations of each item with the full-length validated questionnaire for the same constructs, and (2) assessing the factor structure, internal reliability and measurement invariance across age groups and sex of the composite P-score within a structural equation modelling framework. A secondary aim was to measure validity of the translation process, to justify the pooling of COH-FIT results collected in different languages.

Methods

Dataset

The dataset examined is that from all adult respondents to the COH-FIT survey collected from April 26th, 2020 to May 6th, 2022. Data collection of the full questionnaires after completion of the COH-FIT survey was only conducted between April 26th, 2020 and May 24th, 2020, i.e., until a sufficient number of participants answered these additional questions, in order to reduce the burden and time requirement for the subsequent COH-FIT participants. The validated questionnaires were deliberately placed at the end of the survey in order not to alter the survey's structure before them after removal of these items due to completion of the validation effort. Validation scales were translated into several languages with responses distributed as follows: Hungarian (25%), Italian (20%), Greek (15%), Danish (8%), Thai (8%), English (4%), French (4%), German (4%), Spanish (4%), Japanese (2%), Dutch (1%), Polish (1%), Portugal Portuguese (1%), Russian (1%), Turkish (1%), Romanian (<1%), Traditional Chinese (<1%), Arabic (<1%), Brazilian Portuguese (<1%), Czech (<1%). In addition, the entire WHO-5 questionnaire(Topp et al., 2015) (co-primary outcome with the P-score) was also administered in Bangladeshi, Simplified Chinese, Farsi, Korean, Rumantsch Grischun, Serbian, Swedish, Urdu and Xhosa.

Data screening, languages and missing data

Prior to the main analyses, initial data were screened through computation of minimum and maximum values for each variable to identify out-of-range values. Furthermore, a visual inspection of histograms was conducted to assess data distributions and identify obvious univariate outliers. In addition, for participants who completed ≥80% of the scale items, missing domain item data were imputed using multivariate chained equations. Otherwise participant data for that domain were excluded from further analysis. For COH-FIT domains with a low number of items (typically 1-2 items), domain scores were not imputed if missing.

Representativeness of the validation sample

To assess representativeness of the subsample that additionally completed the full-length validation questionnaires, to the wider survey sample, we compared demographic characteristics based on the following: sex, age, ethnicity, education and employment status. If any sizeable/material imbalance emerged between the validation subsample and the whole data sample, validation cases were weighted to achieve representativeness.

COH-FIT items and concurrent validity

Concurrent validity was assessed by computing Pearson's correlations for each of the candidate COH-FIT domain scores with an established and validated full-length measure of the same construct as follows: (1) COH-FIT anxiety domain score with GAD-7(Spitzer et al., 2006), (2) COH-FIT depression, (3) sleep, and (4) concentration domain score with PHQ-9(Kroenke et al., 2001), (5) COH-FIT post-traumatic symptoms domain score with (PCL-5)(Blevins et al., 2015), (6) COH-FIT obsessive-compulsive symptoms domain score with the Brief Obsessive Compulsive Scale (BOCS)(Bejerot et al., 2014), (7) COH-FIT bipolar disorder symptoms domain score with the Altman Self-Rating Mania Scale (ASRM)(Altman et al., 1997), (8) COH-FIT stress domain score with the WHO-5 wellbeing scale(Topp et al., 2015), and (9) COH-FIT psychotic symptoms domain score with the Prodromal Questionnaire-16 (PQ-16)(Ising et al., 2012). We selected the WHO-5 as the validated questionnaire to test concurrent validity of the COH-FIT stress domain score, given the large overlap between the two concepts (i.e. stress as opposite of well-being)(Heitor Dos Santos et al., 2018), and in light of the strong association between the WHO-5 and several stress signs and symptoms(Feicht et al., 2013). Only COH-FIT domains with moderate correlations $\geq$0.50 with their respective validated full-length questionnaires were considered as acceptable to be included as a component in the composite P-score. Additionally, we calculated the correlations of each individual COH-FIT item within the same domain (e.g. COH-FIT anxiety items 1 and 2) with its corresponding validation scale (e.g. GAD-7 anxiety score) to identify any poorly performing individual COH-FIT items. Any items with a correlation <.20 were not included in the scoring of that domain.

As the upper limit of a test-criterion correlation is dependent upon the reliability of the criterion, the nature of the construct and the degree of similarity of constructs across test, and criterion measures(Kline, 2000), we only automatically excluded COH-FIT domains or items from any analysis where correlations were <0.2, but where correlations were 0.2-0.5 we considered the centrality of that item to the main analysis before deciding whether or not to exclude. The scoring of COH-FIT domains and each corresponding validation scale is provided in Supplementary Table 1. To assess whether concurrent validity for each COH-FIT domain was still evident across different language translations of the COH-FIT items, Pearson's correlations for every language with at least 100 valid responses were computed for all domains and plotted graphically for all COH-FIT domains (Supplementary Table 2, supplementary Figure 2). If any correlations were notably lower for a particular language within a domain, we will consider excluding data for this domain for the affected translation in further projects using global and local data.

P-score definition and internal validation

One of the two COH-FIT co-primary outcomes is a composite psychopathology measure (P-score) representing a multidimensional measure of symptoms of different psychopathologic domains (the other COH-FIT co-primary outcome is a re-scaled WHO-5 questionnaire), with all COH-FIT items and the WHO-5 being rated on a 0 -100 scale. Only COH-FIT domains with at least moderate correlations of $r \geq .50$ with their respective validated full-length questionnaires were considered as acceptable to be included as a component of the composite P-score.

The P-score assessment underwent an internal validation procedure. First, to identify an initial P-score factor structure, we conducted exploratory factor analysis (EFA) on a testing set after randomly dividing the data into approximately evenly split testing and validation subsamples. Factors were extracted from an initial pool of all items belonging to a COH-FIT domain using ordinary least squares EFA, with oblique rotation (oblimin) used, given our expectation of correlated factors. Horn's parallel analysis(Horn, 1965) was performed to determine the number of factors to retain, based on the number of ranked eigenvalues from the data that exceeded the upper 95th percentile of ranked eigenvalues generated from factor analysis of 500 simulated uncorrelated datasets.(Glorfeld, 1995). We used Horn's Parallel analysis, as it is a more objective method than the often used method of visually identifying a "break-point". Parallel analysis determines the number of factors based on how many produce Eigenvalues that lie outside of the 95% confidence intervals of those that would be expected to arise purely by chance, and has been shown to more reliably estimate the 'true' number of factors (Horn, 1965). A rotated item loading >0.45 was considered acceptable for the COH-FIT item, indicating that an item belonged to a factor(Tabachnick and Fidell, 2013). Second, we performed confirmatory factor analysis (CFA) on the validation set, assessing the fit of a hierarchical model using the domain-specific factors identified by the EFA as well as of an additional general

psychopathologic (P) domain modelled as a second-order factor. This general second-order factor was added to evaluate the legitimacy of computing a single composite P-score in further analyses. A substantial loading of the P-domain onto all subfactors (minimum ≥0.45), and an adequate model fit with a general pattern of coherent high factor loadings (minimum ≥0.45), would support the creation of a composite P score. To demonstrate adequacy of model fit indices should be close to the following standard cut-offs of comparative fit index (CFI) >0.95, root mean square error of approximation (RMSEA) <.06, standardized root mean square residual (SRMR) <.08(Hu and Bentler, 1999). We did not use the chi-square test to assess significance of model fit, as even trivial deviations of a user-specified model from a fully saturated model tend to be significant when sample sizes are large (here n>20,000). Overall and individual internal domain reliabilities were estimated with coefficient ω within the CFA framework as well as the traditional coefficient α, given that α can sometimes misestimate true reliability(Raykov, 2001).

P-score measurement invariance

To assess equivalence of P-score measurement across males and females and age groups (18-39, 40-64, 65+ yrs.), multiple-groups CFA was performed. Measurement invariance was tested in a hierarchical manner, assessing adequacy of model fit with the following increasingly restrictive equality constraints:(Horn and McArdle, 1992; Vandenberg and Lance, 2000) configural ('weak') measurement invariance (equal model specifications for each subgroup) and metric ('strong') invariance (equal factor loadings across groups). We also examined intercept invariance (equal intercepts across groups). As limitations of the chi-square test in large samples are also applicable to multi-group CFA, the CFI was used as the primary indicator of measurement invariance. Data simulations have demonstrated that an absolute change in CFI <0.002 (ΔCFI < 0.002) indicates that deviations from perfect group equivalence are practically trivial(Meade et al., 2008).

All analyses were conducted in *R*(R Foundation for Statistical Computing, Vienna, n.d.) using the *MICE* (van Buuren and Groothuis-Oudshoorn, 2011), *ggplot2*(Wickham, 2016), *psych*(Revelle and Revelle, 2015) and *lavaan*(Rosseel, 2012) packages.

Results

Data screening

Up to May 6th, 2022, *N* = 153,876 adults consented to participate in the survey. During the early period of data collection, a smaller subsample was additionally asked to complete a set of full-length validation questionnaires. After approximately 15% (N = 22,456) of the entire sample had provided responses to the validation questionnaires, these were removed from COH-FIT to reduce participant burden. A smaller subsample was available for the PQ-16 scale, which was added at a slightly later stage of the validation process (N = 16,518). A larger sample was available for the WHO-5, as this

scale was also one of the two co-primary outcomes in the main survey (and therefore a complete dataset was available) (Supplementary Table 1).

Only a very small percentage of missing item data were evident and imputed according to the procedure described above, with the vast majority of participants (ranging from 98.0% of participants for the SBQ to 99.9% for the ASRM) completing at least 80% of the total number of items for each questionnaire. Completion rates >80% of all items was similarly high for all COH-FIT domains (ranging from 97.1% for COH-FIT post-traumatic domain to 99.0% for COH-FIT anxiety domain). Data screening found no out-of-range values. Histograms of full-length validation scales and COH-FIT domains are shown in Supplementary Figure 1 and reveal some negative skew in several validation items, as would be expected, given the non-clinical population. However, given the high sample size and that the skew was generally in the same direction for a COH-FIT validation scale domain, we did not attempt to normalise data, as the sampling distribution from which confidence intervals are derived should exhibit normality, given the tenets of the central limit theorem(Lohr, 2010).

Sample demographics and validation sample representativeness

Demographic characteristics of both the entire survey sample and those who completed the validation sample are provided in Table 1. To assess representativeness of the validation sample to the wider survey population, demographic characteristics for each sample were reported, suggesting that the validation subset provides a broadly representative sample of the survey population.

Concurrent validity

Across all COH-FIT items, only one item exhibited a correlation coefficient < 0.20, namely the "mood swings" item from the COH-FIT bipolar disorder symptom domain ($r = 0.05$ with the ASRM). This item was therefore not included in the scoring of the COH-FIT bipolar disorder symptom domain.

Figure 1 and Supplementary Table 1 show the correlation between COH-FIT domains and relative validation questionnaires. Overall, all but the COH-FIT bipolar disorder and OCD symptom domains met our threshold of $r \geq 0.50$. As can be seen in Supplementary figure 2-3, the associations between COH-FIT ratings and external scale scores were generally highly consistent across language translations for each domain (see Supplementary Table 2 for detailed reporting of correlation coefficients).

P-score

As the OCD and bipolar disorder symptom COH-FIT domains did not meet our criteria for acceptable concurrent validity, these were not considered as candidate P-Score domains and therefore excluded

from exploratory factor analysis (EFA). Complete data across remaining domains was available for N = 103,529, and this data set was randomly divided into a testing (N =51,629) and validation (N=51,900) subsets.

Horn's parallel analysis(Horn, 1965) for the remaining COH-FIT domains (anxiety, depression, PTSD, psychosis, sleep, focus and stress) was conducted, on the testing subset, with results showing that five factors were retained (Supplementary Figure 4). COH-FIT item descriptions text, details on how to compute the COH-FIT P-score, and results of the EFA with five extracted factors are presented in the pattern matrix in Table 2 and show all item-factor loadings >0.45 with no complex loadings. Correlations between factors were largely moderate (mean r = 0.58, range = 0.27 to .77), and factor structure was largely consistent with the individual COH-FIT domains, with sleep, focus and stress loading together on a distinct "psychophysiologic" factor.

CFA on the validation set using a model, which included the 5 factors identified by EFA along with a single general factor, suggested a good model fit, with all fit indices satisfying the predefined thresholds, i.e., *CFI* = 0.98, *RMSEA* = 0.053, *SRMR* = 0.028. High indicator-factor loadings for domain-specific factors (0.66 to 0.94) were also observed, with high loadings of the P-score factor onto the five domain-specific factors (Figure 2), consistent with the existence of a general common factor and supporting the aggregation of all domain scores to a general P-score. Unstandardized loadings, standard errors and p-values for the CFA are presented in Supplementary Table 3.

Overall and individual internal scale reliabilities, estimated through ω and α coefficients, are shown in Supplementary Table 5 and suggest good reliability for the five domain-specific factors and excellent reliability for the composite P-score factor, with values above 0.70-0.80 (most commonly used as thresholds for good reliability)(Lance et al., 2006).

P-score measurement invariance

Adequate model fit of the general factor model continued to be demonstrated when CFA was conducted separately in male (CFI = 0.97, RMSEA = 0.067, SRMR = 0.045) and female (CFI = 0.97, RMSEA = 0.061, SRMR = 0.044) subsamples, as well as across age groups of 18-39 years (CFI = 0.96, RMSEA = 0.067, SRMR = 0.051), 40-64 years (CFI = 0.97, RMSEA = 0.065, SRMR = 0.046) and 65+ years (CFI = 0.98, RMSEA = 0.056, SRMR = 0.040).

Factor loadings for each of these subgroups are shown in Supplementary Table 4 and appear to be generally closely equivalent across groups.

Measurement invariance tests results are shown in Supplementary Table 5. All Δ*CFI*s < 0.002 for sex suggest little appreciable degradation in model fit with each increasingly restrictive constraint. For age, some degradation in model fit was shown for factor loading invariance (*CFI* < 0.002) , and intercept invariance (*CFI* = 0.004). Nevertheless, absolute model fit indices retained acceptable fit for all invariance models for both age and sex groups.

Discussion

Results of this validation study show that the selected individual COH-FIT items are valid, providing reliable estimates of individual mental health domains assessed with lengthier validated scales. The selected and implemented COH-FIT items that survive the stricter validity threshold compose a P-score that is internally valid, representing one second order factor (P-score), and five first order factors (anxiety, depression, and post-traumatic, psychotic, and psychophysiologic symptoms).

The translation process of the COH-FIT study proved to be solid, and responses recorded in different COH-FIT study languages can be reliably put together within or across countries.

Several reasons might explain why the bipolar and obsessive-compulsive disorder symptom domains did not meet our validity threshold. Regarding bipolar symptoms, manic symptoms have a low prevalence even in patients with bipolar disorder. For instance, over a follow-up of 11 years, only 4% and 0.4% of subjects with bipolar disorder type I and II, respectively, showed clinically relevant manic symptoms(Fiedorowicz et al., 2009). Hence, these symptoms might be too infrequent to be captured. Moreover, ASRM's specificity is not high with regards to mild manic or hypomanic symptoms, which are expected to be more frequent in the general population(Fiedorowicz et al., 2019). We chose ASRM as there is currently no comparison of psychometric performances of questionnaires to assess manic symptoms in the general population (i.e. no gold standard). Regarding obsessive-compulsive symptoms, the COVID-19 pandemic has certainly elevated the intensity and frequency of thoughts about and, even, preoccupations with contamination, infection, cleanness, and related behaviours to prevent and avoid COVID-19 infection. Such thoughts and behaviours, which are functional, adaptive, and physiologic during infection times, might have altered the psychometric properties of the full-length validated questionnaire, as well as of the corresponding abbreviated OCD COH-FIT item domain. A systematic review focusing on OCD during the COVID-19 pandemic reported a discrepancy in frequency of OCD between in-person versus online studies, with the latter reporting higher rates of OCD, possibly indicating poorer psychometric performance of established tools to screen for OCD during the COVID-19 pandemic and/or using questionnaires(Guzick et al., 2021). Moreover, a more recent scoping review described that obsessive-compulsive symptoms in the general population were associated with trait compulsivity and pandemic-related-stress(Grant et al., 2022), which can confound symptom assessment and impact the validity of the COH-FIT domain extracted from the entire BOCS. Whether the lack of validity of OCD self-ratings affects the full validated questionnaire during time of a pandemic goes beyond the scope of this work, which mainly aims to validate COH-FIT questionnaire and not to test validity of full-length established valid questionnaires for which clinical interviews to diagnose manifest OCD would be needed.

Results are methodologically relevant, as they show that few specific items can be extracted from validated questionnaires for many relevant psychopathology domains and still reliably measure the

whole domain that the complete questionnaire is measuring. The complete PHQ-9 is certainly superior in providing a more detailed and specific symptomatic profile compared with two COH-FIT items. Few items provide a less granular insight of individual symptoms of depressed mood, for instance. However, the PHQ-9 still cannot provide measures of syndromal DSM-5 defined disorders, still being a self-report measure. Thus, unless each of the nine symptoms assessed with PHQ-9 needs to be assessed to test a specific hypothesis, fewer items might be a good trade-off between minimum required validity and broadness of an overall mental health assessment performed in future surveys. Furthermore, results of this study clearly show that multi-language translations of online surveys, scaling them up from local to global surveys is feasible and valid. Beyond broadening the target population internationally, having a multi-language survey within a given country is also of value for inclusivity and representativeness. Selection bias invariably affects online surveys, for instance just because of their online nature (not everybody has access/is familiar with internet), and in particular if convenience sampling is adopted. Selection bias can be counterbalanced by also collecting nationally representative samples via polling agencies, but still, if the survey is available in one language only, those not fluent in the country's main language will be left behind, will not answer, or will provide unreliable responses.

In this study, we applied the gold-standard psychometric procedure for internal and external validation of a questionnaire, namely exploratory factor analysis, and confirmatory factor analysis, measured internal consistency, and tested concurrent/external validation with validated questionnaires. Similar methodologically strict approaches have been used in some but not all (online) surveys conducted during COVID-19. However, most of these scales focused on only one psychopathological domain, or specifically focused particularly on COVID-19, making these questionnaires very specific for the current pandemic setting, but less applicable to future public health crises or infection times. Examples of such new scales developed during the COVID-19 pandemic are the "fear of COVID-19 scale"(Martínez-Lorca et al., 2020), the "COVID-19 anxiety scale"(Chandu et al., 2020), the "Coronavirus Anxiety Scale"(Lee et al., 2020), COVID-19 Public Stigma Scale(Nochaiwong et al., 2021), COVID-19 Exposure and Family Impact Scale(Kazak et al., 2021), COVID-19 Protective Motivation Scale(Cornejo et al., 2021), and a questionnaire on fear of COVID-19 vaccination in the general population(Kumari et al., 2021), to mention a few. Among these aforementioned and many more examples of COVID-19 focused questionnaires that underwent psychometric validation, one stands out as broader and measuring multiple mental health domains, namely the COVID-19 Pandemic Mental Health Questionnaire (CoPaQ)(Rek et al., 2021). CoPaQ measures COVID-19-specific stressor impact, mental health impact, positive coping, institutional and political trust, and conspiracy beliefs, actually going beyond mental health. However, important differences exist between CoPaQ and COH-FIT. First, within the mental health domain, CoPaQ considered PTSD symptoms, sleep disturbance (both part of the broader COH-FIT P-score), and also

substance abuse. COH-FIT deliberately avoided measures of externalizing behaviour in the P-score, a-priori assuming that to properly assess such a domain, in-person assessment and collateral information would be crucial. Results of the methodologically sound CoPaQ validation analyses show that substance abuse poorly correlated with mental health-validated questionnaires (correlation coefficients all below 0.3), confirming that including externalizing symptom- or behavior-related proxy measures in online surveys can be problematic. These results are not surprising, given the evidence of low reliability of questionnaires for the measurement of externalizing behaviors(Dirks and Boyle, 2010). Second, authors did not extract CoPaQ items from validated questionnaires, but created COVID-19-specific questions. Notwithstanding the high specificity and value of CoPaQ during COVID-19, such a methodological approach resulted in overall low correlations of CoPaQ mental health domains with validated questionnaires (all correlation coefficient below 0.5), limiting the applicability of CoPaQ outside of the COVID-19 pandemic.

Results of this present study need to be interpreted within its strengths and limitations. Strengths include that our approach at least mitigated some of the most frequent biases of online surveys(Lin et al., 2021), and subjective reported experiences(Bull et al., 2019), including selection bias (by including representative samples, and by comparing characteristics of validated questionnaire completers versus non completers showing no material demographic differences), short data collection duration (continuous data collection-currently over two years), small sample size (including >150,000 adult study participants as of May 2022, including 22,456 adults who also completed the validated questionnaires), and by testing and verifying internal and concurrent validity of the selected items and questionnaires (across languages). A limitation is that, for the P-score, we only considered internalizing symptoms and thought disorder, but did not include externalizing symptoms or behaviors. As stated above, this decision was deliberate (see design papers)(Solmi et al., 2022c, 2022b), and accounts for poor validity of measures of externalizing behaviors in the context of surveys(Dirks and Boyle, 2010). Additional limitations are inherent to its cross-sectional design. However, participants were at least asked to retrospectively recall key assessed outcomes at the time just before the pandemic started, in order to compare outcomes before and during the pandemic. While this methodology is vulnerable to recall bias, we at least mitigated against the big risk of large attrition in prospective cohort studies. Another limitation for the comparability with other work is that the P-score that we validated in this study parallels the P-factor construct, yet there are some differences. First, the P-factor encompasses externalizing symptoms, P-score does not. As mentioned above, this decision was deliberate. Beyond limited external validity of surveys measures of externalizing behaviours, the current pandemic introduces a global quasi-experimental scenario, with a large drop in several externalizing behaviours, including crime(Ejrnæs and Scherg, 2022; Nivette et al., 2021), and heterogeneous changes of substance (ab)use, and related intoxications, which vary across settings with different lockdown policies. For instance, in the US, where milder lockdown

restrictions were implemented, intoxication and overdose emergency presentations increased(Chandran et al., 2021), while in other settings with stricter lockdown policies substance use did not increase, or decreased(Armstrong et al., 2022; Mason et al., 2022). Hence, while COH-FIT did collect data on substance use and/or domestic violence, we opted not to consider those outcomes as part of the P-score, which is why we did not validate related COH-FIT items. Secondly, COH-FIT models the P-score as an outcome, with a specific quantifiable score, psychometric properties, conceiving it as a measure of mental health, rather than only as a vulnerability factor measuring additional risk of developing or worsening mental disorders(Caspi and Moffitt, 2018). We acknowledge that, to test the P-score as a transdiagnostic vulnerability factor for different mental disorders, future studies will be needed, which should account for structured a-priori transdiagnosticity assessment frameworks(Fusar-Poli et al., 2019), and appropriate prognostic or prediction study designs. Such studies should be cohort studies measuring the P-score at baseline and following-up in participants over time, measure prognostic accuracy, discrimination performance, each in development, internal, and external validation samples(Meehan et al., 2022; Salazar de Pablo et al., 2021).

Despite these limitations, in COH-FIT we were able to develop and validate across multiple languages and in a reasonably large sample a P-score consisting of multiple clinically relevant internalizing symptom domains that can be useful for research during the current COVID pandemic and other crisis situations affecting mental well-being and functioning.

In conclusion, COH-FIT is a valid tool to measure clinically relevant domains of mental health during infections times, which is available in 30 languages and provides a measure of overall mental health via a composite P-score. These results are relevant for the use of the P-score in forthcoming analyses and publications from the COH-FIT study but also for other questionnaire studies in the future. Whether the P-score reflects current psychopathology, or also increased vulnerability for mental disorders, or both, needs to be clarified in additional longitudinal studies.

Author disclosures

Funding Statement

All the institutions and funding agencies are listed in supplementary Table 6. COH-FIT PIs and collaborators have applied/are actively applying for several national and international grants to cover expenses related to the coordination of the study, website, nationally representative samples, advertisement of the study, and future dissemination of study findings.

Conflict of Interest Statement

Conflict of interest statements of all authors are detailed in Supplementary table 7.

Author Contributions Statement

TT wrote the statistical analysis plan designed the statistical analysis plan and conducted the analysis of this work. CUC, MS, TT wrote the first draft. All authors read, contributed to and approved the final version of the manuscript. For the overall COH-FIT project, MS, CUC wrote the study protocol. MS, CUC, TT, SC, FL, QR, AI, ED CUC, MS, AA, AE, DV conducted a preliminary review of the available publications and ongoing registered studies. All authors contributed to the final version of the COH-FIT survey and are involved in disseminating the COH-FIT survey link and collecting the data and designing and preparing research reports on national data. All local researchers contributed to and approved translations of the COH-FIT survey in their respective language. CUC, MS, AE, ED, TT, FL, AK had access to the global raw data on participation results.

Acknowledgements

All authors thank all respondents who took the survey so far, funding agencies and all professional and scientific national and international associations supporting or endorsing the COH-FIT project.

References

Altman, E.G., Hedeker, D., Peterson, J.L., Davis, J.M., 1997. The Altman Self-Rating Mania Scale.

Armstrong, M.J., Cantor, N., Smith, B.T., Jesseman, R., Hobin, E., Myran, D.T., 2022. Interrupted time series analysis of Canadian legal cannabis sales during the COVID-19 pandemic. Drug Alcohol Rev. 10.1111/dar.13465. https://doi.org/10.1111/dar.13465

Bejerot, S., Edman, G., Anckarsäter, H., Berglund, G., Gillberg, C., Hofvander, B., Humble, M.B., Mörtberg, E., Rastam, M., Stahlberg, O., Frisén, L., 2014. The Brief Obsessive-Compulsive Scale (BOCS): A self-report scale for OCD and obsessive-compulsive related disorders. Nord. J. Psychiatry 68, 549–559. https://doi.org/10.3109/08039488.2014.884631

Blevins, C.A., Weathers, F.W., Davis, M.T., Witte, T.K., Domino, J.L., 2015. The Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5): Development and Initial Psychometric Evaluation. J. Trauma. Stress 28, 489–498. https://doi.org/10.1002/jts.22059

Bull, C., Byrnes, J., Hettiarachchi, R., Downes, M., 2019. A systematic review of the validity and reliability of patient-reported experience measures. Health Serv. Res. 54, 1023–1035. https://doi.org/10.1111/1475-6773.13187

Caspi, A., Houts, R.M., Belsky, D.W., Goldman-Mellor, S.J., Harrington, H., Israel, S., Meier, M.H., Ramrakha, S., Shalev, I., Poulton, R., Moffitt, T.E., 2014. The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? Clin. Psychol. Sci. a J. Assoc. Psychol. Sci. 2, 119–137. https://doi.org/10.1177/2167702613497473

Caspi, A., Moffitt, T.E., 2018. All for One and One for All: Mental Disorders in One Dimension. Am. J. Psychiatry 175, 831–844. https://doi.org/10.1176/appi.ajp.2018.17121383

Chandran, K., Mazer-Amirshahi, M., Shankar, N., Desale, S., Nelson, L., Mete, M., 2021. Impact of COVID-19 pandemic on emergency department substance use screens and overdose presentations. Am. J. Emerg. Med. 50, 472–476. https://doi.org/10.1016/j.ajem.2021.08.058

Chandu, V.C., Pachava, S., Vadapalli, V., Marella, Y., 2020. Development and Initial Validation of the COVID-19 Anxiety Scale. Indian J. Public Health 64, S201–S204. https://doi.org/10.4103/IJPH.IJPH_492_20

Chen, J., Zhang, S.X., Yin, A., Yáñez, J.A., 2022. Mental health symptoms during the COVID-19 pandemic in developing countries: A systematic review and meta-analysis. J. Glob. Health 12, 5011. https://doi.org/10.7189/jogh.12.05011

Cornejo, B., Vela, B., Vilca, L.W., Vallejos, M., Huancahuire-Vega, S., 2021. Psychometric Properties of the COVID-19 Protective Motivation Scale in Peruvians During the Health Emergency. J. Prim. Care Community Health 12, 21501327211051936–21501327211051936. https://doi.org/10.1177/21501327211051935

Dirks, M.A., Boyle, M.H., 2010. The comparability of mother-report structured interviews and checklists for the quantification of youth externalizing symptoms. J. Child Psychol. Psychiatry

51, 1040–1049. https://doi.org/https://doi.org/10.1111/j.1469-7610.2010.02244.x
Dong, E., Du, H., Gardner, L., 2020. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis. https://doi.org/10.1016/S1473-3099(20)30120-1
Dragioti, E., Li, H., Tsitsas, G., Lee, K.H., Choi, J., Kim, J., Choi, Y.J., Tsamakis, K., Estradé, A., Agorastos, A., Vancampfort, D., Tsiptsios, D., Thompson, T., Mosina, A., Vakadaris, G., Fusar-Poli, P., Carvalho, A.F., Correll, C.U., Han, Y.J., Park, S., Il Shin, J., Solmi, M., 2021. A large-scale meta-analytic atlas of mental health problems prevalence during the COVID-19 early pandemic. J. Med. Virol. n/a. https://doi.org/https://doi.org/10.1002/jmv.27549
Dragioti, E., Tsartsalis, D., Mentis, M., Mantzoukas, S., Gouva, M., 2022. Impact of the COVID-19 pandemic on the mental health of hospital staff: An umbrella review of 44 meta-analyses. Int. J. Nurs. Stud. 131, 104272. https://doi.org/10.1016/j.ijnurstu.2022.104272
Ejrnæs, A., Scherg, R.H., 2022. Nightlife activity and crime: The impact of COVID-19 related nightlife restrictions on violent crime. J. Crim. Justice 79, 101884. https://doi.org/10.1016/j.jcrimjus.2022.101884
Feicht, T., Wittmann, M., Jose, G., Mock, A., von Hirschhausen, E., Esch, T., 2013. Evaluation of a seven-week web-based happiness training to improve psychological well-being, reduce stress, and enhance mindfulness and flourishing: a randomized controlled occupational health study. Evid. Based. Complement. Alternat. Med. 2013, 676953. https://doi.org/10.1155/2013/676953
Fiedorowicz, J.G., Persons, J.E., Assari, S., Ostacher, M.J., Zandi, P., Wang, P.W., Thase, M.E., Frye, M.A., Coryell, W., 2019. Depressive symptoms carry an increased risk for suicidal ideation and behavior in bipolar disorder without any additional contribution of mixed symptoms. J. Affect. Disord. 246, 775–782. https://doi.org/https://doi.org/10.1016/j.jad.2018.12.057
Fiedorowicz, J.G., Solomon, D.A., Endicott, J., Leon, A.C., Li, C., Rice, J.P., Coryell, W.H., 2009. Manic/hypomanic symptom burden and cardiovascular mortality in bipolar disorder. Psychosom. Med. 71, 598–606. https://doi.org/10.1097/PSY.0b013e3181acee26
Fusar-Poli, P., Solmi, M., Brondino, N., Davies, C., Chae, C., Politi, P., Borgwardt, S., Lawrie, S.M., Parnas, J., McGuire, P., 2019. Transdiagnostic psychiatry: a systematic review. World Psychiatry 18, 192–207. https://doi.org/10.1002/wps.20631
Glorfeld, L.W., 1995. An Improvement on Horn's Parallel Analysis Methodology for Selecting the Correct Number of Factors to Retain. Educ. Psychol. Meas. 55, 377–393. https://doi.org/10.1177/0013164495055003002
Grant, J.E., Drummond, L., Nicholson, T.R., Fagan, H., Baldwin, D.S., Fineberg, N.A., Chamberlain, S.R., 2022. Obsessive-compulsive symptoms and the Covid-19 pandemic: A rapid scoping review. Neurosci. Biobehav. Rev. 132, 1086–1098. https://doi.org/10.1016/j.neubiorev.2021.10.039
Guzick, A.G., Candelari, A., Wiese, A.D., Schneider, S.C., Goodman, W.K., Storch, E.A., 2021.

Obsessive-Compulsive Disorder During the COVID-19 Pandemic: a Systematic Review. Curr. Psychiatry Rep. 23, 71. https://doi.org/10.1007/s11920-021-01284-2

Heitor Dos Santos, M.J., Moreira, S., Carreiras, J., Cooper, C., Smeed, M., Reis, M. de F., Pereira Miguel, J., 2018. Portuguese version of a stress and well-being evaluation tool (ASSET)at the workplace: validation of the psychometric properties. BMJ Open 8, e018401–e018401. https://doi.org/10.1136/bmjopen-2017-018401

Horn, J.L., 1965. A rationale and test for the number of factors in factor analysis. Psychometrika 30, 179–185. https://doi.org/10.1007/BF02289447

Horn, J.L., McArdle, J.J., 1992. A practical and theoretical guide to measurement invariance in aging research. Exp. Aging Res. 18, 117–144. https://doi.org/10.1080/03610739208253916

Hu, L., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct. Equ. Model. A Multidiscip. J. 6, 1–55. https://doi.org/10.1080/10705519909540118

Ising, H.K., Veling, W., Loewy, R.L., Rietveld, M.W., Rietdijk, J., Dragt, S., Klaassen, R.M.C., Nieman, D.H., Wunderink, L., Linszen, D.H., van der Gaag, M., 2012. The validity of the 16-item version of the Prodromal Questionnaire (PQ-16) to screen for ultra high risk of developing psychosis in the general help-seeking population. Schizophr. Bull. 38, 1288–1296. https://doi.org/10.1093/schbul/sbs068

Kazak, A.E., Alderfer, M., Enlow, P.T., Lewis, A.M., Vega, G., Barakat, L., Kassam-Adams, N., Pai, A., Canter, K.S., Hildenbrand, A.K., McDonnell, G.A., Price, J., Schultz, C., Sood, E., Phan, T.-L., 2021. COVID-19 Exposure and Family Impact Scales: Factor Structure and Initial Psychometrics. J. Pediatr. Psychol. 46, 504–513. https://doi.org/10.1093/jpepsy/jsab026

Kline, P., 2000. Handbook of Psychological testing, 2nd ed.

Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The PHQ-9: Validity of a brief depression severity measure. J. Gen. Intern. Med. 16, 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kumari, A., Ranjan, P., Chopra, S., Kaur, D., Upadhyay, A.D., Kaur, T., Bhattacharyya, A., Arora, M., Gupta, H., Thrinath, A., Prakash, B., Vikram, N.K., 2021. Development and validation of a questionnaire to assess knowledge, attitude, practices, and concerns regarding COVID-19 vaccination among the general population. Diabetes Metab. Syndr. 15, 919–925. https://doi.org/10.1016/j.dsx.2021.04.004

Lance, C.E., Butts, M.M., Michels, L.C., 2006. The Sources of Four Commonly Reported Cutoff Criteria: What Did They Really Say? Organ. Res. Methods 9, 202–220. https://doi.org/10.1177/1094428105284919

Lee, S.A., Mathis, A.A., Jobe, M.C., Pappalardo, E.A., 2020. Clinically significant fear and anxiety of COVID-19: A psychometric examination of the Coronavirus Anxiety Scale. Psychiatry Res.

290, 113112. https://doi.org/10.1016/j.psychres.2020.113112
Leung, C.M.C., Ho, M.K., Bharwani, A.A., Cogo-Moreira, H., Wang, Y., Chow, M.S.C., Fan, X., Galea, S., Leung, G.M., Ni, M.Y., 2022. Mental disorders following COVID-19 and other epidemics: a systematic review and meta-analysis. Transl. Psychiatry 12, 205. https://doi.org/10.1038/s41398-022-01946-6
Lin, Y.H., Chen, C.Y., Wu, S.I., 2021. Efficiency and quality of data collection among public mental health surveys conducted during the COVID-19 pandemic: Systematic review. J. Med. Internet Res. https://doi.org/10.2196/25118
Lohr, S., 2010. Sampling: Design and Analysis.
Martínez-Lorca, M., Martínez-Lorca, A., Criado-Álvarez, J.J., Armesilla, M.D.C., Latorre, J.M., 2020. The fear of COVID-19 scale: Validation in spanish university students. Psychiatry Res. 293, 113350. https://doi.org/10.1016/j.psychres.2020.113350
Mason, R., Roberts, A., Spaight, R., Shaw, D., Whitley, G.A., Hogue, T.E., Siriwardena, A.N., Rogers, J., Law, G.R., 2022. Ambulance attendance for substance and/or alcohol use in a pandemic: Interrupted time series analysis of incidents. Drug Alcohol Rev. 41, 932–940. https://doi.org/10.1111/dar.13453
Meade, A.W., Johnson, E.C., Braddy, P.W., 2008. Power and Sensitivity of Alternative Fit Indices in Tests of Measurement Invariance. J. Appl. Psychol. 93, 568–592. https://doi.org/10.1037/0021-9010.93.3.568
Meehan, A.J., Lewis, S.J., Fazel, S., Fusar-Poli, P., Steyerberg, E.W., Stahl, D., Danese, A., 2022. Clinical prediction models in psychiatry: a systematic review of two decades of progress and challenges. Mol. Psychiatry 27, 2700–2708. https://doi.org/10.1038/s41380-022-01528-4
Nivette, A.E., Zahnow, R., Aguilar, R., Ahven, A., Amram, S., Ariel, B., Burbano, M.J.A., Astolfi, R., Baier, D., Bark, H.-M., Beijers, J.E.H., Bergman, M., Breetzke, G., Concha-Eastman, I.A., Curtis-Ham, S., Davenport, R., Díaz, C., Fleitas, D., Gerell, M., Jang, K.-H., Kääriäinen, J., Lappi-Seppälä, T., Lim, W.-S., Revilla, R.L., Mazerolle, L., Meško, G., Pereda, N., Peres, M.F.T., Poblete-Cazenave, R., Rose, S., Svensson, R., Trajtenberg, N., van der Lippe, T., Veldkamp, J., Perdomo, C.J.V., Eisner, M.P., 2021. A global analysis of the impact of COVID-19 stay-at-home restrictions on crime. Nat. Hum. Behav. 5, 868–877. https://doi.org/10.1038/s41562-021-01139-z
Nochaiwong, S., Ruengorn, C., Awiphan, R., Kanjanarat, P., Ruanta, Y., Phosuya, C., Boonchieng, W., Nanta, S., Chongruksut, W., Thavorn, K., Wongpakaran, N., Wongpakaran, T., 2021. COVID-19 Public Stigma Scale (COVID-PSS): development, validation, psychometric analysis and interpretation. BMJ Open 11, e048241–e048241. https://doi.org/10.1136/bmjopen-2020-048241
Pan, D., Sze, S., Minhas, J.S., Bangash, M.N., Pareek, N., Divall, P., Williams, C.M., Oggioni, M.R.,

Squire, I.B., Nellums, L.B., Hanif, W., Khunti, K., Pareek, M., 2020. The impact of ethnicity on clinical outcomes in COVID-19: A systematic review. EClinicalMedicine 23. https://doi.org/10.1016/j.eclinm.2020.100404

R Foundation for Statistical Computing, Vienna, A.U. https://www. R. org/., n.d. R Core Team (2019). R: A language and environment for statistical computing.

Raykov, T., 2001. Estimation of congeneric scale reliability using covariance structure analysis with nonlinear constraints. Br. J. Math. Stat. Psychol. 54, 315–323. https://doi.org/https://doi.org/10.1348/000711001159582

Rek, S. V, Bühner, M., Reinhard, M.A., Freeman, D., Keeser, D., Adorjan, K., Falkai, P., Padberg, F., 2021. The COVID-19 Pandemic Mental Health Questionnaire (CoPaQ): psychometric evaluation and compliance with countermeasures in psychiatric inpatients and non-clinical individuals. BMC Psychiatry 21, 426. https://doi.org/10.1186/s12888-021-03425-6

Revelle, W., Revelle, M.W., 2015. Package 'psych.' Compr. R Arch. Netw. 337, 338.

Rosseel, Y., 2012. Lavaan: An R package for structural equation modeling. J. Stat. Softw. 48, 1–36. https://doi.org/10.18637/jss.v048.i02

Salazar de Pablo, G., Studerus, E., Vaquerizo-Serrano, J., Irving, J., Catalan, A., Oliver, D., Baldwin, H., Danese, A., Fazel, S., Steyerberg, E.W., Stahl, D., Fusar-Poli, P., 2021. Implementing Precision Psychiatry: A Systematic Review of Individualized Prediction Models for Clinical Practice. Schizophr. Bull. 47, 284–297. https://doi.org/10.1093/schbul/sbaa120

Solmi, M., Cortese, S., Correll, C.U., 2022a. Editorial Perspective: Challenges of research focusing on child and adolescent mental health during the COVID-19 era: what studies are needed? J. Child Psychol. Psychiatry Allied Discip. 63, 122–125. https://doi.org/10.1111/jcpp.13478

Solmi, M., Cortese, S., Correll, C.U., 2021. Editorial Perspective: Challenges of research focusing on child and adolescent mental health during the COVID-19 era: what studies are needed? J. Child Psychol. Psychiatry Allied Discip. https://doi.org/10.1111/jcpp.13478

Solmi, M., Estradé, A., Thompson, T., Agorastos, A., Radua, J., Cortese, S., Dragioti, E., Leisch, F., Vancampfort, D., Thygesen, L.C., Weiser, M., Correll, C.U., 2022b. Physical and mental health impact of COVID-19 on children, adolescents, and their families: The Collaborative Outcomes study on Health and Functioning during Infection Times - Children and Adolescents (COH-FIT-C&A). J. Affect. Disord. 299, 367–376. https://doi.org/10.1016/j.jad.2021.09.090

Solmi, M., Estradé, A., Thompson, T., Agorastos, A., Radua, J., Cortese, S., Dragioti, E., Leisch, F., Vancampfort, D., Thygesen, L.C., Weiser, M., Correll, C.U., 2022c. The collaborative outcomes study on health and functioning during infection times in adults (COH-FIT-Adults): Design and methods of an international online survey targeting physical and mental health effects of the COVID-19 pandemic. J. Affect. Disord. 299, 393–407. https://doi.org/10.1016/j.jad.2021.07.048

Spitzer, R.L., Kroenke, K., Williams, J.B.W., Löwe, B., 2006. A brief measure for assessing

generalized anxiety disorder: The GAD-7. Arch. Intern. Med. 166, 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Tabachnick, B., Fidell, L., 2013. Using Multivariate Statistics, 6th ed. Boston, MA.

Theberath, M., Bauer, D., Chen, W., Salinas, M., Mohabbat, A.B., Yang, J., Chon, T.Y., Bauer, B.A., Wahner-Roedler, D.L., 2022. Effects of COVID-19 pandemic on mental health of children and adolescents: A systematic review of survey studies. SAGE open Med. 10, 20503121221086710–20503121221086710. https://doi.org/10.1177/20503121221086712

Topp, C.W., Østergaard, S.D., Søndergaard, S., Bech, P., 2015. The WHO-5 Well-Being Index: A Systematic Review of the Literature. Psychother. Psychosom. 84, 167–176. https://doi.org/10.1159/000376585

Treweek, S., Forouhi, N.G., Narayan, K.M.V., Khunti, K., 2020. COVID-19 and ethnicity: who will research results apply to? Lancet. https://doi.org/10.1016/S0140-6736(20)31380-5

UK_Government, 2011. English language skills by ethnicity, 2011 England and Wales.

van Buuren, S., Groothuis-Oudshoorn, K., 2011. mice: Multivariate Imputation by Chained Equations in R. J. Stat. Software; Vol 1, Issue 3 .

Vandenberg, R.J., Lance, C.E., 2000. A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. Organ. Res. Methods 3, 4–70. https://doi.org/10.1177/109442810031002

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis, 2nd ed. TX, USA.

Zhang, S.X., Batra, K., Xu, W., Liu, T., Dong, R.K., Yin, A., Delios, A.Y., Chen, B.Z., Chen, R.Z., Miller, S., Wan, X., Ye, W., Chen, J., 2022. Mental disorder symptoms during the COVID-19 pandemic in Latin America - a systematic review and meta-analysis. Epidemiol. Psychiatr. Sci. 31, e23–e23. https://doi.org/10.1017/S2045796021000767

Figure 1. Pearson's correlation of COH-FIT domain (x-axis) and validation questionnaire (y-axis) measures for each of the COH-FIT domains.

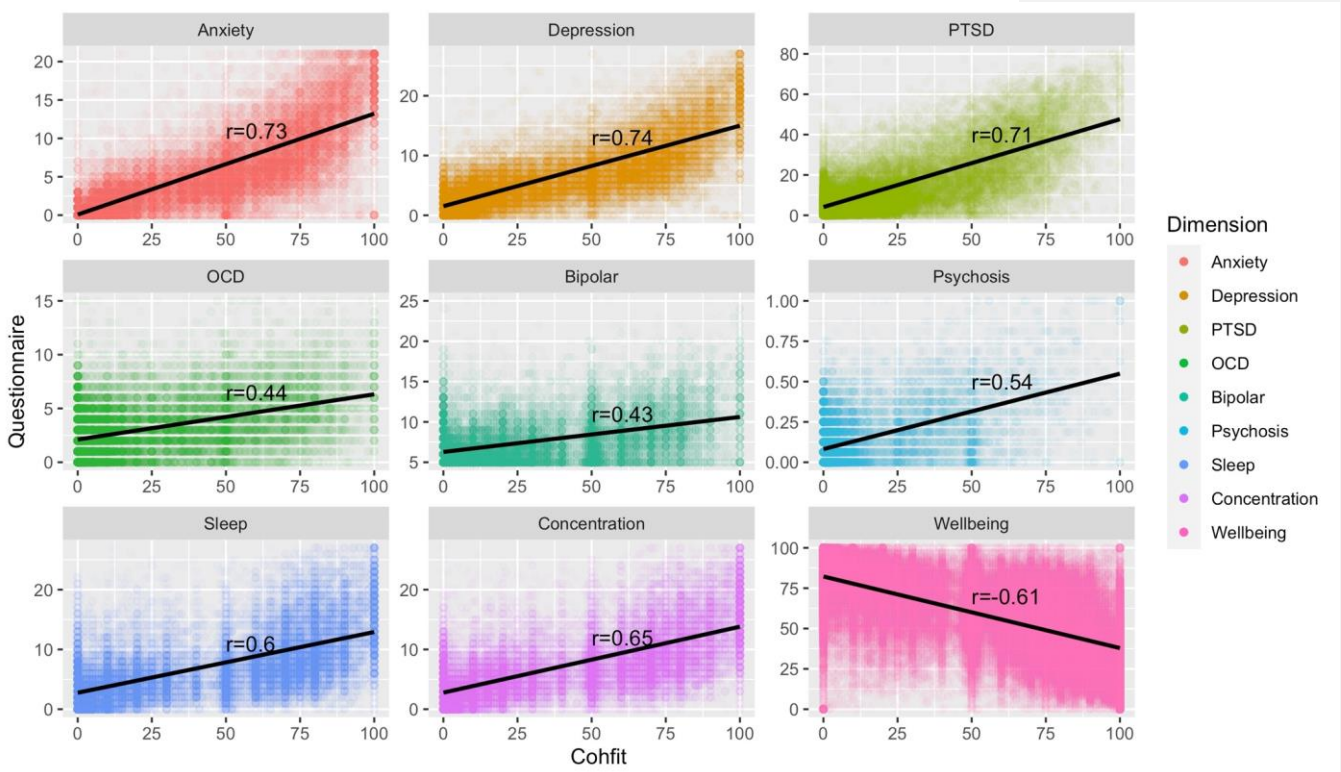

Figure 2. Factor structure of the composite psychopathology P-Score from confirmatory factor analysis.

Commented [MS1]: Trevor please to replace with P-score instead of P-Factor

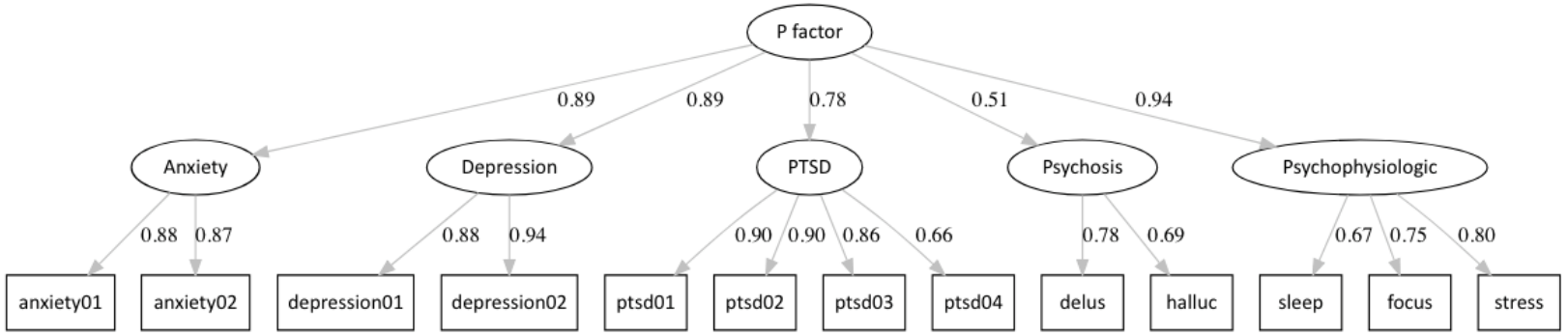

Table 1. Sample demographics

	Validation sample* (N = 22,456)	**Total Survey sample (N = 153,876)**
Gender	Female 69% Male 30% Other or not stated <1%	Female 67% Male 32% Other or not stated <1%
Age	42.5 years (SD = 15.0)	40.8 years (SD = 15.6)
Ethnicity	White 78% Asian 10% Mixed 1% Hispanic 1% African/African-descent <1% Other <1% Not stated 9%	White 69% Asian 20% Mixed 4% Hispanic 3% African/African-descent 3% Other 1% Not stated <1%
Education	None <1% Primary school 2% High school 25% College/university degree 64% PhD 8%	None <1% Primary school 3% High school 29% College/university degree 59% PhD 9%
Job Status	Current paid job 65% No paid job 35%	Current paid job 62% No paid job 38%

*This consisted of those completing the anxiety domain (COH-FIT anxiety and GAD-7). Similar demographic distributions were observed for other domains.

Table 2. COH-FIT items and loading matrix of P-Score in exploratory factor analysis

COH-FIT items*	Anxiety	Depression	PTSD	Psychosis	Psychophysiologic
Anxiety - Over the last two weeks, how often have you been bothered by any of the following problems:					
Feeling nervous, anxious, or on edge? [anxiety01]	.88	.01	-.01	.00	.05
Not being able to stop or control worrying? [anxiety02]	.71	.10	.09	.05	-.02
Depression - Over the last two weeks, how often have you been bothered by any of the following problems:					
Little interest or pleasure in doing things? [depression01]	-.01	.96	-.02	.00	.00
Feeling down, depressed, or hopeless? [depression02]	.14	.69	.08	.01	.06
PTSD - In the last two weeks, how much have you been bothered by any of the following problems, related to a stressful life experience:					
Repeated disturbing memories, thoughts, or images, or dreams of the stressful experience? [ptsd01]	.01	.01	.86	-.04	.05
Suddenly acting or feeling as if the stressful experience was happening again (as if you were reliving it)? [ptsd02]	-.01	-.01	.93	.03	-.04
Avoiding thinking about, or talking about, or having feelings related to, or avoiding engaging in activities or situations that remind you of the stressful experience? [ptsd03]	.00	.02	.84	.00	.01
Being "super alert" or watchful or on guard? [ptsd04]	.17	-.02	.47	.06	.09
Psychosis					
In the last two weeks, how much did you experience any of the following: i) believe that you seem to live through events exactly as they happened before (déjà vu), ii) believe that someone is out to get or harm you on purpose, iii) believe that your thoughts or actions are not your own; iv) see special meanings in advertisements, shop windows, or in the way things are arranged around you, v) believe that you have a very important special purpose or mission in life that others can't understand.? [delusional]	-.05	.03	.12	.60	.04
In the last two weeks, how much did you hear, see, smell, taste or feel things that other cannot? [hallucination]	.03	-.01	-.04	.83	-.01
Psychophysiologic					
In the last two weeks, how much have you experienced sleep problems (difficulty falling or staying asleep, early morning awakening)? [sleep]	.05	.05	.08	.09	.49
How difficult has it been for you to concentrate or focus, in the last two weeks? [focus]	-.11	.13	.03	.05	.70
How stressed have you felt in the last two weeks? [stress]	.14	-.05	.01	-.03	.74
P-score: Compute the mean item score for each of the 5 domains and then sum to create an overall P-score (0 – 500)					

Legend. COH-FIT, Collaborative Outcomes study on Health and Functioning during Infection Times; PTSD, post-traumatic stress disorder; *all COH-FIT items were 0-100 VAS scale.

Table 3. Reliability estimates for general and domain-specific factors

	Anxiety	Depression	PTSD	Psychosis	Psychophysiologic	P-score
omega	.86	.91	.90	.71	.78	.95
alpha	.86	.91	.90	.70	.78	.93