La Medicina del Lavoro

Original article

Med Lav 2023; 114 (1): e2023009
DOI: 10.23749/mdl.v114i1.13918

Psychometric Evaluation of GHQ-12 as a Screening Tool for Psychological Impairment of Healthcare Workers Facing COVID-19 Pandemic

Anna Comotti[1*], Alice Fattori[2], Francesca Greselin[3], Lorenzo Bordini[1], Paolo Brambilla[4,5], Matteo Bonzini[1,2]

[1]Occupational Health Unit, Foundation IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy
[2]Department of Clinical Science and Community Health, University of Milan, Milan, Italy
[3]Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy
[4]Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy
[5]Department of Neurosciences and Mental Health, Foundation IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy

Keywords: General Health Questionnaires (GHQ-12); Occupational stress; Health-Care Workers; Item Response Theory

Abstract

Background: *The General Health Questionnaire (GHQ) is a widely used tool in clinical and research settings due to its brevity and easy administration. Researchers often adopt a dichotomous measurement method, considering a total score above or below a certain threshold, leading to an extreme simplification of the gathered data and, therefore, the loss of clinical details. In a multistep evaluation study aimed at assessing health care workers' mental health during the COVID-19 pandemic, GHQ-12 proved to be the most effective tool to detect psychological distress compared to other scales. These results deepened the understanding of GHQ-12 properties through a statistical study focusing on items' properties and characteristics.* **Methods:** *GHQ-12 responses were analyzed using Item Response Theory (IRT), a suitable method for scale assessment. Instead of considering the single overall score, in which each item accounts equally, it focuses on individual items' characteristics. Moreover, IRT models were applied combined with the latent class (LC) analysis, aiming to determine subgroups of individuals according to their level of psychological distress.* **Results:** *GHQ-12 was administered to 990 healthcare workers, and responses were scored using the binary method (0-0-1-1). We applied the two-parameter logistic (2-PL) model, finding that the items showed different ways of responses and features. The latent class analysis classified subjects into three sub-groups according to their responses to GHQ-12 only: 47% of individuals with general well-being, 38% expressing signs of discomfort without severity, and 15% of subjects with a high level of impairment. This result almost reproduces the subjects' classification obtained after administering the six questionnaires of the study protocol.* **Conclusions:** *Accurate statistical techniques and a deep understanding of the latent factors underlying the GHQ-12 resulted in more effective usage of such a psychometric questionnaire – i.e., a more refined gathering of data and significant time and resource efficiency. We underlined the need to maximize the extraction of data from questionnaires and the necessity of them being less lengthy and repetitive.*

Received 09.11.2022 - Accepted 23.01.23
Corresponding Author: Anna Comotti, Occupational Health Unit, Foundation IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy; E-mail: anna.comotti@policlinico.mi.it

1. INTRODUCTION

The General Health Questionnaire (GHQ) aims to provide information about an individual's mental well-being by identifying distressing symptoms [1]. Its shorter version (GHQ 12-item) has become one of the most widely used scales for assessing psychological distress and short-term changes in mental health, and its popularity can be mainly attributable to its brevity and easy administration [2].

GHQ-12 has shown strong psychometric properties and it is recommended as screening tool to detect common mental disorders as depressive, anxiety and somatic disorders [3, 4].

Several analyses explored its characteristics, especially the factor structure, mostly identifying a two-factor solution through Confirmatory Factor Analysis (CFA) techniques. The two factors commonly proposed were the Depression/Anxiety construct (related to the emotional component of psychological distress) and the Social Dysfunction construct (related to the social functioning component of the individual experiencing the distress) [5]. Other authors proposed a three-dimensional solution, comprising anxiety (4-item), social dysfunction (6-item), and loss of confidence (2-item) dimensions [6].

The use of GHQ-12 to measure the mental health status in healthcare populations is frequent, and several recent contributions gave examples of its application in analyzing psychological well-being during the pandemic [7-14]. In all these cases, the screening of the psychological status through GHQ-12 was determined according to its total score. The scores typically used are the binary scale (0-0-1-1) and the 4-point Likert-type scale (0-1-2-3). Responses to all items are summed up to a total score ranging from 0 to 12 (binary scale) or 0 to 36 (Likert scale), with higher scores indicating more severe impairment. A score above a specific cut-off (3/4 for bimodal and 13/14 for the Likert scale) indicates psychological distress and suggests further investigation for potential mental disorders [15].

A possible difference in items contribution can be lost through such a measurement method, in which each item counts the same. Indeed, every single item may have a different weight, expressing different severity of the psychological impairment measured by the test.

We proposed to analyze GHQ-12 data through Item Response Theory (IRT)-based methods as they provide more details about individual items. IRT is a specific statistical model for evaluating questionnaires, and it is a more suitable tool than the usual methodologies based on Classical Test Theory (CTT), whose use is still prevalent in the psychometric field. The strength of such a technique lies in its focus on items rather than individual scores, while in the CTT, the evaluation of test properties and item characteristics is not included.

From a statistical point of view, if the results of a test are reported as a single score, it is implicitly assumed that all the items are measuring the same trait equally therefore losing the complexity of underlying traits in psychological testing. IRT allows to evaluate individual-level distress and to describe the performances of the items on the questionnaire simultaneously may providing a better clinical insight on symptoms detected and associations with potential underlying mental disorders.

To our knowledge, few authors proposed analyzing the GHQ-12 scale via an IRT approach. In some cases, IRT was applied as a suitable tool to determine the factor structure of the scale [16, 17], and more recently through the multidimensional version of IRT [18]. For instance, the IRT approach used by Smith et al. [16] explored the fact that item phrasing, item variance and levels of respondents' distress affect the factor structure observed for the GHQ-12 and may perhaps explain why different factor structures of the instrument have been found in different populations. Other uses of IRT on GHQ-12 regard computerized adaptive testing [19] or Mokken analysis [20].

However, no studies have studied the performance of GHQ-12 with IRT in an occupational setting during a pandemic. Therefore, this work aims to perform an IRT-based analysis on GHQ-12, investigating the methodological and clinical benefits of such an approach.

2. METHODS

2.1 Population Study

We conducted a multistep epidemiological study within occupational health surveillance to systematically assess healthcare workers' mental well-being

during the COVID-19 pandemic in a large Hospital in Milan (Italy).

GHQ-12 was administered, jointly with the Impact of Event Scale (IES-R; post-traumatic distress, [21]) and General Anxiety Disorder Scale (GAD-7; anxiety, [22]) questionnaires, to assess the psychological impact of the pandemic and to identify possible signs of impairment, further investigated through psycho-diagnostic questionnaires and specialist evaluation.

The steps of such assessment were fully presented in a previous study [23]: for each worker, the psychological well-being was screened in three steps. The first-level questionnaire collected several personal information and data from three tests (i.e., GHQ-12, IES-R, and GAD-7). Workers who scored above the cut-off in at least one scale were further investigated by the second-level questionnaire composed of psycho-diagnostic scales to assess depressive symptoms (Patient Health Questionnaire-9; [24]), dissociative symptoms (Dissociative Experiences Scale –II; [25]) and other psychological symptoms (Symptoms Checklist-90 [26]). If the second level showed psychological impairments, an individual specialized treatment with a psychiatrist and psychologist (third-level) was offered.

In this framework, GHQ-12 (binary version) proved to be an effective screening tool, deserving a deeper investigation [27].

The occupational medicine unit, where workers underwent the periodical health surveillance already prescribed by the current Italian legislation, proposed the study protocol to all workers since July 2020. By July 2021, 990 subjects out of a total population of 1,610 had been enrolled in the study. The participation rate was 62%. In detail, 220 (13%) workers did not answer our calls or were unavailable and 400 (25%) refused to participate.

Participants were predominantly female (70%) with a mean age of 45 years (sd=11); nurses (42.5%) was the most prominent job category, followed by physicians (23.5%), administrative staff (12%), health assistants (6.5%) and other roles (16%). Four hundred and forty-six (45%) participants had the experience of working in a COVID-19 area: 25% were working with COVID-19 patients during data collection, and 20% had worked in a COVID-19 department before enrollment.

Six hundred and twenty-seven workers (63%) did not show signs of psychological impairment; 363 (37%) presented signs of psychological impairment at the first screening level (i.e., with scores above the cut-off in at least one scale among GHQ-12, IES-, R and GAD-7) and underwent the second level assessment. Among these, 231 (67%; 23% of the total sample) scored above the cut-off in at least one scale among PHQ-9, DES, and SCL-90. As a result, we were able to classify participants into three sub-groups, according to their scorings: a group with no evidence of psychological distress after first-level screening (Group 1, N=627), workers who expressed distress without severe symptoms (Group 2, N=132), and subjects who expressed signs of impairment and received psychological and/or psychiatric support (Group 3, N=231).

Out of the 363 subjects who showed psychological impairment at first-level screening, almost all (91%) scored above the cut-off (equal to 4) of GHQ-12, while about half of them over-passed the cut-off of IES-R and GAD-7 scales (53% and 56% respectively). This result suggested that GHQ-12 could determine the transition to the second level more effectively than the other scales.

Results obtained from the analysis of risk factors for psychological impairment were presented in detail in a previous paper [27].

2.2 Item Response Theory (IRT)

The basic assumption of IRT models is that a person's interactions with test items can be represented according to probabilistic relations, containing a single parameter to describe the individual's characteristics θ. The power of IRT is that it estimates item characteristics through some item parameters, which permit the calculation of the expected score at the item level (e.g., probability of 1 or correct answer if responses are binary or dichotomous) and at the test level. In addition, the person's latent trait θ_i for an individual *i* is also estimated, considering specific item characteristics and how the person answers to each item.

We applied the so-called two parameters logistic (2-PL), suitable for binary data, which uses two parameters to describe each item j, corresponding to its "difficulty" and its "discrimination". The item

difficulty represents the level of latent trait for which one has a 50% probability of responding 'correctly' (or 1) to that item. In other words, if $\theta_i=\beta_j$, then $P(Y_{ij}=1)=0.5$. If $\theta_i>\beta_j$, then $P(Y_{ij}=1)>0.5$ and if $\theta_i<\beta_j$, then $P(Y_{ij}=1)<0.5$. The discriminating parameter for item j, λ_j, estimates the capacity of the item to distinguish between subjects with different latent trait levels.

Two plots are typically employed in the IRT framework to visualize the analysis results. Item Characteristic Curves (ICCs) show the probability of answering 1 to an item at varying levels of the latent trait, specifying how well an item discriminates between respondents at various levels of the latent trait. The "easier" items functions are on the left side of the plot, in the lower regions of the latent trait scale, while the more "difficult" items are on the right (in our case, they are the items that underlie more severe impairment in mental health). The discrimination parameter represents the slope, which refers to how well the item response options discriminate (or differentiate) between subjects with high and low latent trait levels.

IICs show how well and precisely each item measures the latent trait at various attribute levels. Item Information measures the strength of the relationship between an item and the latent trait. Some items may provide more information at low levels of the attribute, while others may provide more information at higher levels of the attribute.

We applied the IRT model in one of its discrete versions, based on the so-called Latent Class (LC) analysis [28, 29], whose assumption is that the population under study is composed of homogeneous classes of individuals who have very similar unobservable characteristics [30, 31]. Data were collected through a computerized database generated by REDCap [32], which was subsequently analyzed by R software [33].

3. RESULTS

Based on the dichotomous scored version of GHQ-12, we calculated Cronbach's α equal to 0.87, indicating good internal consistency. The mean score was 3.31 (SD=3.45), with 37% of subjects scoring above the cut-off equal to 4 (indicating a general level of psychological distress). Table 1 presents the twelve questionnaire items with the distribution of the respective answers.

Item parameters estimation (difficulty and discrimination) of the 2-PL model are reported and graphically represented in Figure 1, which shows the ICCs.

Item 5 (followed by Item 7) has the lowest threshold parameter, while Item 3 ("feeling useless") and Item 11 ("thinking of yourself as a worthless person") are the most 'difficult' ones. Item 5 has a location (or difficulty) parameter equal to 0, meaning that a person with a latent trait θ at level 0 has the same probability of answering 0 or 1 to Item 5. On the contrary, a level of latent trait θ equal to 1.73 is needed for having an equal probability of answering less than/same as usual or more/much more than usual to Item 11. Concerning the discrimination parameter, Item 12 (feeling reasonably happy, all things considered) has the highest value, much greater than the others. On the other hand, Item 3 has the lowest discrimination parameter. The item provides sample information about differences across individuals when discrimination is high. Item 5 and Item 7 have the leftmost lines represented in the plot, while the curve of Item 3, which is also the less steep, is plotted on the right. The curve of Item 12 is indeed the steepest one. In the IICs plot in Figure 2, the curve of Item 12 gives much information around a value

Table 1. GHQ-12 answers distribution.

	0	**1**
Item 1 - Able to concentrate	75%	25%
Item 2 - Loss of sleep over worry	64%	36%
Item 3 - Playing a useful part	86%	14%
Item 4 - Capable of making decisions	85%	15%
Item 5 - Felt constantly under strain	49%	51%
Item 6 - Could not overcome difficulties	78%	22%
Item 7 - Able to enjoy day-to-day activities	55%	45%
Item 8 - Able to face problems	79%	21%
Item 9 - Feeling unhappy and depressed	68%	32%
Item 10 - Losing confidence	83%	17%
Item 11 - Thinking of self as worthless	92%	8%
Item 12 - Feeling reasonably happy	73%	27%

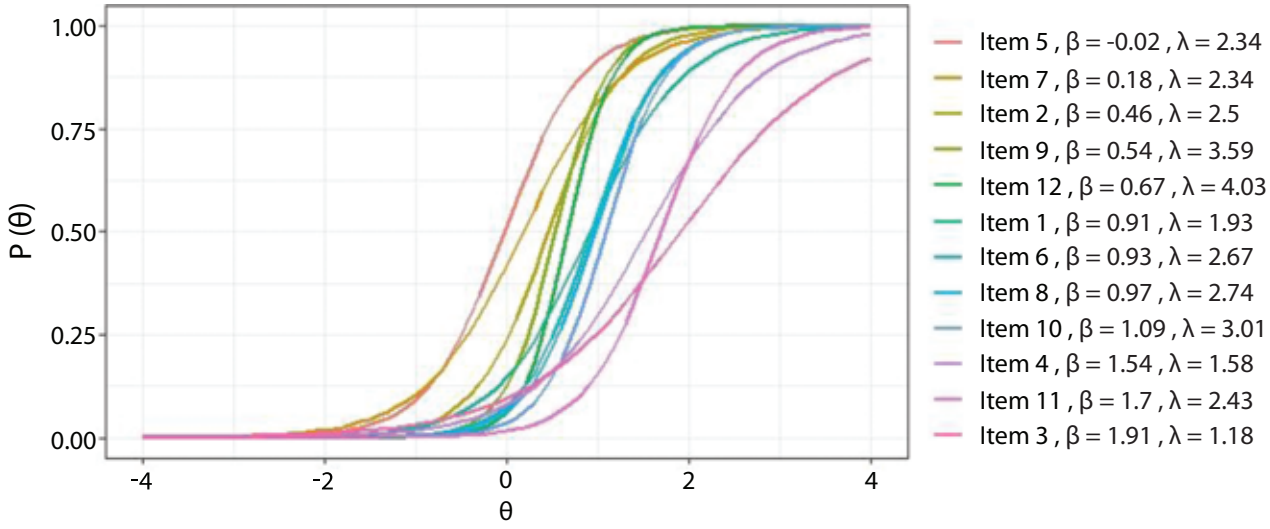

Figure 1. Item Characteristic Curves for GHQ-12 questionnaire.

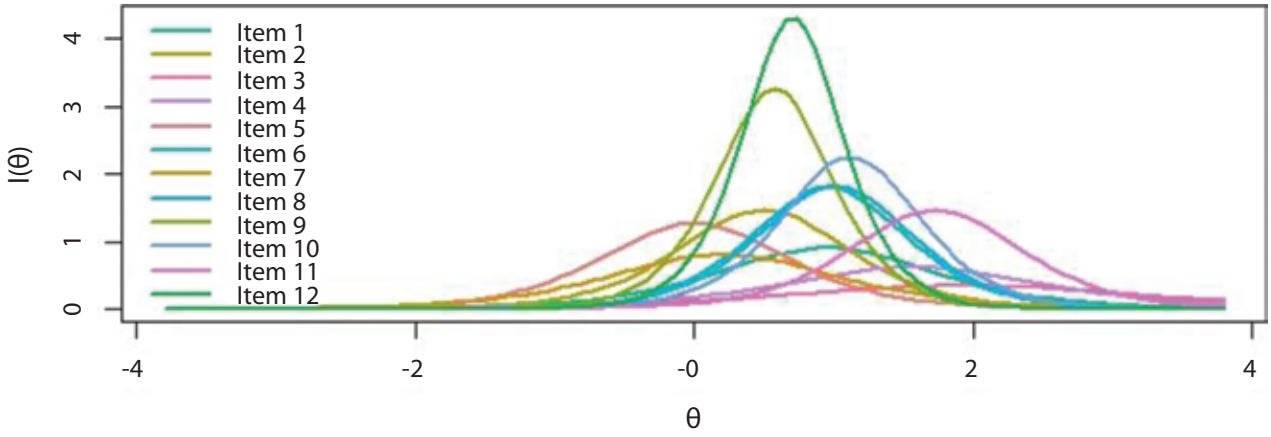

Figure 2. Item Information Curves for GHQ-12 questionnaire.

of θ between 0 and 1, while Item 3 (with the lowest discrimination parameter) gives more or less the same (low) information over a broader range.

The latent class model reaches the best fit (calculated through BIC) with three latent classes. The weights and levels of the latent trait for each dimension and latent class are in Table 2. The latent model estimates three values for the θ finding support points equal to -3.3 (low level of distress) with weight equal to 0.47, a medium level (around 0) for 38% of subjects, and a higher level of distress (θ=1.15) with weight 0.15.

We compared the three latent classes with the three groups resulting from questionnaire scorings (Group 1, Group 2, Group 3) in Table 2. Almost all (97%) subjects belonging to Class 1 did not undergo the second-level screening, i.e., they did not express any sign of discomfort through GHQ-12, IES-R, and GAD-7. On the contrary, most (85%) of those assigned to Class 3 needed psychological support, while only one-third of Class 2 required psychological therapy.

Table 3 shows the percentage of answers equal to 1 for each item, according to the latent class.

Table 2. Latent class model results and percentage of subjects by multistep evaluation.

	θ	%	Group 1 (no distress)	Group 2 (psychological distress)	Group 3 (psychological impairment)
Class 1	-3.3	47%	97%	1%	2%
Class 2	-0.8	38%	44%	20%	36%
Class 3	1.2	15%	1%	14%	85%

Table 3. Probability of answering equally to 1 by class.

	ITEM											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Class 1	3%	4%	4%	2%	17%	1%	16%	1%	1%	-	-	-
Class 2	31%	52%	17%	17%	76%	24%	64%	21%	45%	15%	6%	33%
Class 3	76%	92%	38%	50%	96%	81%	90%	81%	95%	76%	41%	97%

Participants assigned to Class 1 answered 0 for almost all the items, with the highest percentages of the answer 1 occurring for Items 5 and 7 (but much less than in the general distribution). On the contrary, considering such items (Items 5 and 7), almost everyone who belongs to Class 3 answered 0. In addition, for the group with more severe signs of psychological distress, percentages of answer 1 were much higher (than the general distribution) up to the items found to be the most "difficult" (Item 3 and Item 11). For the second class, the distribution was more balanced, and more than half of the participants answered 1 only to Item 2, Item 5, and Item 7.

4. Discussion

The GHQ-12 is frequently used among different settings and populations, and its assessment methods adopt predominantly a dichotomous scoring, which may contribute to lose potential differences in items contribution; this rationale motivated us to a psychometric analysis to better clarify the methodological and clinical quality of this tool. The analysis was carried out within a study aimed to evaluate psychological well-being of healthcare workers in a large Hospital in Milan (Italy) facing COVID-19 pandemic.

In our scenario, the IRT was a suitable tool for scale assessment. Instead of considering the single overall score, in which each item accounts equally, the item-based analysis produced interesting results by identifying specific items able to detect psychological impairment effectively.

Such considerations are similar to those obtained in previous analyses on GHQ-12 based on IRT methods in other frameworks. For example, the approach of Smith and colleagues [18], within the multi-dimensionality assessment of GHQ-12, showed how the use of the summated scores for the GHQ-12 could potentially lead to an incorrect assessment of patients' psychiatric morbidity.

The focus on items characteristics allowed us to deeply investigate how the mental health status was captured by GHQ-12 in our population, identifying different levels of severity (given by item difficulty) and quantifying the impact each item had on the measurement of general distress. We further specify that participants were healthcare workers involved in a disruptive pandemic, which imposed them unprecedented and heavy workloads coupled with lack pf preparation to cope with such demands. In such circumstances, questions about utility, capacity to make decisions, loss of trust and confidence showed peculiar responses, affecting in different way the psychological wellbeing; feeling useless (Item 3) and thinking of yourself as a worthless person (Item 11) caused more severe impairments than, for instance, feeling constantly under strain (Item 5) or being unable to enjoy day-to-day activities (Item 7).

The analysis on the item response patterns also allowed the classification of subjects according to different impairment levels: the first class with almost all responses equal to 0 (subjects without distress), the second class where percentages of responses equal to 1 were high only for Item 3, Item 5 and Item 9 (subjects with psychological distress) and the third class with huge percentages of responses equal to 1 (subjects with psychological impairment).

Such a classification agreed with previous results obtained by administering several other psychological questionnaires. Indeed, through an item-based latent class analysis, we could determine the screening outcome without considering the other questionnaires, previously part of the first-level evaluation (IES-R, GAD-7).

IRT was a helpful tool for identifying clinically meaningful subgroups in our population, recognizing distinct patient profiles, and tailoring effective interventions, whose importance was already underlined in previous works [34].

Thus, subjects' classification based only on responses to GHQ-12 could potentially simplify workers evaluation. Results show that one step of the evaluation (i.e. second-level) is redundant and may be skipped. According to symptoms' severity, immediate access to specialist evaluation can be planned for those with psychological impairment (Class 3), without testing them through second-level scales; subjects with less severity, i.e. psychological distress (Class 2), will be instead be monitored with a check evaluation after a certain period of time.

Our study is prone to potential biases as self-selection of respondents [35]. We managed to minimize that risk grounding our investigation on the occupational physician health surveillance, obtaining a very high participation rate and minimizing the risk of untrue or uncompleted answers.

We know that our results cannot be generalized, and neither are they comparable with results obtained in different scenarios. The pandemic's consequences directly affected our population, and this exceptional situation should be carefully considered. However, in terms of psychological assessment, our results agree with findings obtained in similar populations of healthcare workers who expressed high levels of GHQ-12 during the COVID-19 pandemic in Italy (e.g., more than 50% above cut-off equal to 3 in Del Piccolo et al. [10]) and in other countries (e.g., 39% of subjects above cut-off equal to 3 in Dai et al. [8]).

5. Conclusions

The GHQ-12 is commonly analyzed and interpreted according to CTT rules and we decided to complement it by performing an analysis based on IRT. As outlined in our work, drawing on the strengths of IRT as an alternative to CTT analyses supported the development of rigorous measures and valuable interpretations. It was possible to classify the degree of severity of psychological impairment by administering only GHQ-12 questionnaire and according to response patterns, focusing on the way in answering each question more than the scored obtained as a sum of "positive responses".

In light of these results, our approach may suggest simplifying the multistep protocol for evaluating mental health in occupational settings, recommending using GHQ-12 as a single measurement tool to be the most effective. Such a method may also meet the need for resources and time reduction when conducting studies and assessments involving workers.

Through such analysis, we gave an example of the utility of IRT in psychometric studies conducted among workers populations. The application of appropriate methodological tools to support the interpretation of questionnaires could sensibly discover their potential in simplifying the screening framework and saving one of the most important workers' resources: their time. Even in questionnaire-based epidemiological studies, in many cases, *less is more*.

Funding: The full cost of the study was covered by the Foundation IRCCS Ca' Granda Ospedale Maggiore Policlinico (internal funding) and by a charitable donation to the University of Milan by the "Fondazione Romeo ed Enrica Invernizzi".

Institutional review board statement: The study was approved by the Hospital Ethical Committee (Milan Area 2 Ethical Committee, n.652_2020 of July 21, 2020) and was conducted in compliance with all local legal and regulatory requirements, Good Clinical Practice, the International

Conference on Harmonisation document and the Declaration of Helsinki.

Informed consent statement: The participation was voluntary; each subject read and signed an extended informed-consent to participate in the study.

Acknowledgments: The Authors thank all workers who agreed to participate and undergo our multistep evaluation for giving us their precious time. Further, authors thank Ezio Belleri, Laura Chiappa, Dario Laquintana and all the IRCCS Cà Granda Fondazione Hospital managing team for their support in different phases of study design and implementation; Antonio Minolfi and Stefania Pazzi from the Pavia (Italy) Consorzio di Bioingenieria e Informatica Medica (CBIM) for the data collection software and their technical support; Chiara Moltrasio, MariaLuna Sandri and Francesca Maddalena for their technical support in contacting study subjects and arranging workers' interview.

Declaration of interest: The authors declare no conflict of interest.

References

1. Baldwin JA. The Detection of Psychiatric Illness by Questionnaire. By DP Goldberg. Maudsley Monograph No. 21. Oxford University Press. 1972.
2. Goldberg DP. General Health Questionnaire-12. *Aust J Psychol*. 1978. Doi: https://doi.org/10.1037/t00297-000
3. Goldberg D. GHQ and psychiatric case. *Br J Psychiatry*. 1979;134:446-447.
4. Goldberg DP. The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychol. Med.* 1997;27(1):191-197.
5. Andrich D, van Schoubroeck L. The General Health Questionnaire: a psychometric analysis using latent trait theory. *Psychol Med.* 1989;19(2):469-485. Doi:10.1017/s0033291700012502
6. Graetz B. Multidimensional properties of the General Health Questionnaire. *Soc Psychiatry Psychiatr Epidemiol.* 1991;26(3):132-138. Doi:10.1007/bf00782952
7. Marton G, Vergani L, Mazzocco K, Garassino MC, Pravettoni G. 2020s Heroes Are Not Fearless: The Impact of the COVID-19 Pandemic on Wellbeing and Emotions of Italian Health Care Workers During Italy Phase 1. *Front Psychol*. 2020;11:588762. Published 2020 Oct 15. Doi:10.3389/fpsyg.2020.588762
8. Dai Y, Hu G, Xiong H, Qiu H, Yuan X. Psychological impact of the coronavirus disease 2019 (COVID-19) outbreak on healthcare workers in China. *bioRxiv.* Published online 2020. Doi:10.1101/2020.03.03.20030874
9. Shoja E, Aghamohammadi V, Bazyar H, et al. Covid-19 effects on the workload of Iranian healthcare workers. *BMC Public Health*. 2020;20(1):1636. Published 2020 Nov 2. Doi:10.1186/s12889-020-09743-w
10. Del Piccolo L, Donisi V, Raffaelli R, et al. The Psychological Impact of COVID-19 on Healthcare Providers in Obstetrics: A Cross-Sectional Survey Study. *Front Psychol*. 2021;12:632999. Published 2021 Apr 9. Doi:10.3389/fpsyg.2021.632999
11. Feng J, Xu J, Xu S, et al. Psychological Impact During the First Outbreak of COVID-19 on Frontline Health Care Workers in Shanghai. *Front Public Health*. 2021;9:646780. Published 2021 May 17. Doi:10.3389/fpubh.2021.646780
12. Mascayano F, van der Ven E, Moro MF, et al. The impact of the COVID-19 pandemic on the mental health of healthcare workers: study protocol for the COVID-19 HEalth caRe wOrkErS (HEROES) study. *Soc Psychiatry Psychiatr Epidemiol*. 2022;57:633-645. Doi:10.1007/s00127-021-02211-9
13. Mediavilla R, Fernández-Jiménez E, Martínez-Alés G, et al. Role of access to personal protective equipment, treatment prioritization decisions, and changes in job functions on health workers' mental health outcomes during the initial outbreak of the COVID-19 pandemic. *J Affect Disord.* 2021;295:405-409. Doi:10.1016/j.jad.2021.08.059
14. Tanaka K, Tahara M, Mashizume Y, Takahashi K. Effects of Lifestyle Changes on the Mental Health of Healthcare Workers with Different Sense of Coherence Levels in the Era of COVID-19 Pandemic. *Int J Environ Res Public Health*. 2021;18(6):2801. Published 2021 Mar 10. Doi:10.3390/ijerph18062801
15. Goldberg DP, Gater R, Sartorius N, et al. The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychol Med*. 1997;27(1):191-197. Doi:10.1017/s0033291796004242
16. Smith AB, Fallowfield LJ, Stark DP, Velikova G, Jenkins V. A Rasch and confirmatory factor analysis of the general health questionnaire (GHQ)-12. *Health Qual Life Outcomes*. 2010;8:45. Published 2010 Apr 30. Doi:10.1186/1477-7525-8-45
17. Böhnke JR, Croudace TJ. Calibrating well-being, quality of life and common mental disorder items: psychometric epidemiology in public mental health research. *Br J Psychiatry.* 2016;209(2):162-168. Doi:10.1192/bjp.bp.115.165530
18. Nouri F, Feizi A, Roohafza H, Sadeghi M, Sarrafzadegan N. How different domains of quality of life are associated with latent dimensions of mental health measured by GHQ-12. *Health Qual Life Outcomes.* 2021;19(1):255. Published 2021 Nov 14. Doi:10.1186/s12955-021-01892-9
19. Stochl J, Böhnke JR, Pickett KE, Croudace TJ. An evaluation of computerized adaptive testing for general psychological distress: combining GHQ-12 and Affectometer-2 in an item bank for public mental health

research. *BMC Med Res Methodol.* 2016;16:58. Published 2016 May 20. Doi:10.1186/s12874-016-0158-7
20. Stochl J, Jones PB, Croudace TJ. Mokken scale analysis of mental health and well-being questionnaire item responses: a non-parametric IRT method in empirical research for applied health researchers. *BMC Med Res Methodol.* 2012;12:74. Published 2012 Jun 11. Doi:10.1186/1471-2288-12-74
21. Weiss DS, Marmar CR. Impact of event scale--revised. *PsycTESTS Dataset*. 1997. Doi:10.1037/t12199-000
22. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med.* 2006;166(10):1092-1097. Doi:10.1001/archinte.166.10.1092
23. Fattori A, Cantù F, Comotti A, et al. Hospital workers mental health during the COVID-19 pandemic: methods of data collection and characteristics of study sample in a university hospital in Milan (Italy). *BMC Med Res Methodol.* 2021;21(1):163. Published 2021 Aug 10. Doi:10.1186/s12874-021-01355-1
24. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med.* 2001;16(9):606-613. Doi:10.1046/j.1525-1497.2001.016009606.x
25. Bernstein EM, Putnam FW. Development, reliability, and validity of a dissociation scale. *J Nerv Ment Dis.* 1986;174(12):727-735. Doi:10.1097/00005053-198612000-00004
26. Derogatis LR, Lipman RS, Covi L. SCL-90: an outpatient psychiatric rating scale--preliminary report. *Psychopharmacol Bull.* 1973;9(1):13-28.
27. Bonzini M, Comotti A, Fattori A, et al. One Year Facing COVID. Systematic Evaluation of Risk Factors Associated With Mental Distress Among Hospital Workers in Italy. *Front Psychiatry*. 2022;13:834753. Published 2022 Mar 10. Doi:10.3389/fpsyt.2022.834753
28. Lazarsfeld PF. Latent Structure Analysis. Boston: Houghton Mifflin Company; 1968.
29. Goodman LA. Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*. 1974;61(2):215-231. Doi:10.1093/biomet/61.2.215
30. Bartolucci F. A class of multidimensional IRT models for testing unidimensionality and clustering items. *Psychometrika*. 2007;72(2):141-157. Doi:10.1007/s11336-005-1376-9
31. Bartolucci F, Bacci S, Gnaldi M. Statistical Analysis of Questionnaires: A Unified Approach Based on R and Stata. Chapman & Hall Book/CRC; 2015.
32. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international community of software platform partners. *J Biomed Inform.* 2019;95:103208. Doi:10.1016/j.jbi.2019.103208
33. R Core Team. A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2020). Available online at: https://www.R-project.org/
34. Prenovost KM, Fihn SD, Maciejewski ML, Nelson K, Vijan S, Rosland AM. Using item response theory with health system data to identify latent groups of patients with multiple health conditions. *PLoS One.* 2018;13(11):e0206915. Published 2018 Nov 26. Doi:10.1371/journal.pone.0206915
35. Nguyen LH, Drew DA, Graham MS, et al. Risk of COVID-19 among front-line health-care workers and the general community: a prospective cohort study. *Lancet Public Health*. 2020;5(9):e475-e483. Doi:10.1016/S2468-2667(20)30164-X