Measuring Population Mental Health

OECD

Measuring Population Mental Health

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Note by all the European Union Member States of the OECD and the European Union
The Republic of Cyprus is recognised by all members of the United Nations with the exception of Türkiye. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

Please cite this publication as:
OECD (2023), *Measuring Population Mental Health*, OECD Publishing, Paris, https://doi.org/10.1787/5171eef8-en.

ISBN 978-92-64-63593-7 (print)
ISBN 978-92-64-62576-1 (pdf)
ISBN 978-92-64-78244-0 (HTML)
ISBN 978-92-64-74029-7 (epub)

Photo credits: Cover © cienpies /123rf.com, additional design by Sonia Primot (Statistics and Data Directorate)

Corrigenda to publications may be found on line at: www.oecd.org/about/publishing/corrigenda.htm.

© OECD 2023

The use of this work, whether digital or print, is governed by the Terms and Conditions to be found at https://www.oecd.org/termsandconditions.

Foreword

Mental health is increasingly being recognised as a public priority in today's societies and economies. Understanding and mapping both its positive and negative dimensions is key to informing a number of public policies and actions by the private sector and civil society. This report aims to support national statistical offices and other data producers in collecting high-quality measures of population mental health outcomes in a more frequent, consistent and internationally harmonised manner. It documents existing measurement practice across OECD countries, discusses the advantages and limitations of available measurement tools, and recommends priority measures (for both mental ill-health and positive mental states) to adopt in household, social and health surveys. While this report does not present fully-fledged measurement guidelines, it complements past OECD WISE Centre work aiming to advance the statistical agenda on people's well-being, including guidelines on subjective well-being; trust; and the quality of the working environment. This publication is the first of two reports prepared as part of a special assessment of mental health and well-being in the context of the OECD *How's Life?* publication series. The second report applies a well-being lens to population mental health by examining its interlinkages with the different economic, social, environmental and relational dimensions of people's lives as exemplified by the OECD Well-being Framework.

The report was prepared by the OECD WISE Centre. The authoring team consisted of Lara Fleischer (Chapters 1 and 2) and Jessica Mahoney (Chapters 2 and 3). Jessica Mahoney also led the statistical work for this publication. Manuela Grabosch, Muriel Levy and Nikita Arora are gratefully acknowledged for the background research that informed various sections of this report. Lara Fleischer led the project and content editing under the supervision of Carrie Exton, who provided additional editing alongside Marco Mira d'Ercole and Romina Boarini. The report was published under the direction of Romina Boarini. Martine Zaida has provided essential communications support throughout the project. Cassandra Morley prepared and formatted the manuscript for publication, Patrick Hamm copy edited the work and Sonia Primot designed the front cover.

The report has benefitted from valuable comments provided by national delegates to the OECD Committee on Statistics and Statistical Policy and the OECD Health Committee. Their contributions and advice are very gratefully acknowledged.

We are also grateful to many colleagues in the OECD and externally for their help, comments, insights and data, and edits either on draft text or in relation to specific queries. They include, but are not limited to:

- Doron Wijker, Christopher Prinz, Shunta Takino and Emily Hewlett (OECD Employment, Labour and Social Affairs Directorate);
- The members of the Informal Advisory Group supporting this workstream:
  - Adam Coutts (Weatherhead Centre for International Affairs, Harvard University Senior Research Fellow, Magdalene College, University of Cambridge)
  - Alexandra Lazaro, Cath Davies, Jess Rackham, Lily Makurah and Linda Bullivant (Department of Health and Social Care, United Kingdom)

- Amelia Walters (Board Member, Victorian Collaborative Centre for Mental Health and Wellbeing)
- Andrew Steptoe (Professor of Psychology and Epidemiology, Head of the Department of Behavioural Science and Health at University College London)
- Angeline Ferdinand (Research Officer, Centre for Health Policy, Melbourne School of Population and Global Health, University of Melbourne)
- Bernard Jacob (Federal Coordinator of the Belgian Mental Health Care Reforms, Belgian Federal Public Service - Public Health)
- Caroline Cohrdes (Mental Health Unit, Robert Koch Institute)
- Catherine Carty (UNESCO Chair project manager)
- Claire Gibbons (Senior Program Officer, Robert Wood Johnson Foundation)
- Daniel Hugh Chisholm (Programme Manager for Mental Health, World Health Organization Regional Office for Europe)
- David Finch (Assistant Director Healthy Lives Team, The Health Foundation)
- Elisabeth Ng Langdal (Executive Director, Mental Health and Human Rights Info)
- Emma Lawrance and Jessica Newberry Le Vay (Institute of Global Health Innovation, Imperial College London)
- Fredrik Lindencrona (Lead for Strategic Improvement, Swedish Association of Local Authorities and Regions)
- George Ploubidis (Professor of Population Health and Statistics, University College London Social Research Institute)
- Gregory Bratman (Assistant Professor, Director Environment and Well-being Labs, University of Washington)
- Hannes Kröger (Former Head of the SocPsych-MH Unit, German Institute for Economic Research)
- Harold Alan Pincus (Professor and Vice Chair, Department of Psychiatry, Co-Director, Irving Institute for Clinical and Translational Research, Vagelos College of Physicians and Surgeons, Columbia University)
- Hedinn Unnsteinsson (Senior Policy Analyst, Prime Minister's Office of Iceland, Chairperson of the Icelandic Mental Health Alliance)
- Katie Hayes (Senior Policy Analyst, Health Canada)
- Layla McCay (Director, Centre for Urban Design and Mental Health)
- Lene Søvold (Clinical Psychologist, Mental Health Advisor & Researcher)
- Luca Bernardi (Senior Lecturer Politics, University of Liverpool)
- Margaret Walker and Peter Falkai (Executive Director and President, European Psychiatric Association)
- Patricio V Márquez (Senior Associate, Johns Hopkins University Bloomberg School of Public Health, former World Bank Group Lead Public Health Specialist)
- Richard Layard (Co-Director, Community Wellbeing Programme, Centre for Economic Performance, London School of Economics)
- Ronni M. Greenwood (Social-Community Psychologist and Lecturer in the Psychology Department, University of Limerick)
- Sarah Hinde (Acting Assistant Secretary of the Mental Health Data and Evidence Taskforce, Australian Government Department of Health)
- Shekhar Saxena (Professor of the Practice of Global Mental Health at the Harvard T. H. Chan

School of Public Health)
  - Tyler Norris (MDiv, CEO, Well Being Trust)
  - Ziggi Ivan Santini (Researcher, Danish National Institute of Public Health);
- Participants of a project webinar organised for national statistical offices on 28 April 2022:
  - Adriana Pérez Amador and Olinca Páez (INEGI, Mexico)
  - Alessandra Tinto and Lidia Gargiulo (ISTAT, Italy)
  - Camélia St-Denis, Emma Howieson, Isabelle Lévesque, Joie Huggins and Leah Smith (Statistics Canada)
  - Chika Arita (Ministry of Internal Affairs and Communications, Japan)
  - Claudio Peter and Daniela Schuler (Swiss Health Observatory)
  - Emilie Hegelund (Statistics Denmark)
  - Kali Kong (U.S. Bureau of Economic Analysis)
  - Kelly Sabbe (Statistics Belgium)
  - Dace Krievkalne (Central Statistical Bureau, Latvia)
  - Guy Weber (Health Directorate, Luxembourg)
  - Lydia Gisle (Belgian Institute for Health)
  - Matthew Montgomery (Australian Bureau of Statistics)
  - Mónika Földvári (Hungarian Central Statistical Office)
  - Naama Rotem (Central Bearau of Statistics, Israel)
  - Sara Holsbrink (Public Health Agency, Sweden)
  - Rosario Gonzalez Garcia (INE, Spain)
  - Sunhye Choi (Korea Disease Control and Prevention Agency).

Support for this report was provided in part by the Robert Wood Johnson Foundation. The views expressed herein do not necessarily reflect the views of the Foundation.

Table of contents

Foreword 3

Reader's Guide 11

Executive summary 13

1 What is population mental health and why should we measure it? 15
The importance of focusing on population mental health outcomes 16
Concepts of mental health: From illness to wellness 18
Mental ill-health 23
Positive mental health 26
Conclusion 27
References 28
Notes 33

2 Measuring population mental health: Tools and current country practice 35
Which tools are available for measuring population mental health? 37
Which population mental health data are OECD countries already collecting? 45
Conclusion and ways forward 54
Annex 2.A. Mental health survey measures by country 55
Annex 2.B. Details on standardised survey tools to measure mental health 71
References 82
Notes 90

3 Good practices for measuring population mental health in household surveys 91
Statistical quality 95
Data collection 116
Analysis 121
Conclusion 125
References 128
Notes 141

FIGURES

Figure 1.1. The OECD Well-being Framework 17
Figure 1.2. Models of mental health 19
Figure 1.3. The mental health continuum as a river 20
Figure 1.4. The dual continuum of mental health in Australian children 22

Figure 2.1. Trade-off between response burden and accuracy for mental health measurement tools 41
Figure 2.2. Screening tools typically show greater prevalence of major depressive disorder than do structured interviews 43
Figure 2.3. The share of those reporting a diagnosis of a mental health condition is much lower than the share identified as experiencing psychological distress by screening tools 44
Figure 2.4. The majority of OECD countries report measuring population mental health in both health and general social surveys 46
Figure 2.5. Many OECD countries collect mental health data infrequently, with over half reporting four-to-ten-year lags between survey rounds 47
Figure 2.6. All OECD countries reported collecting data on mental ill-health and positive mental health, with the latter mostly focused on life evaluation 49
Figure 2.7. Screening tools and questions about experience of symptoms and disorders are the most common mental ill-health tools reported by countries 50
Figure 2.8. Screening tools capturing general psychological distress and symptoms of depression are more commonly used than those for symptoms of anxiety or other disorders 51
Figure 2.9. Countries do capture anxiety data, but often with non-standardised measures 52
Figure 2.10. Affect data are more commonly collected than eudaimonic data, but OECD countries are not aligned in the tools used to collect data on positive mental health beyond life satisfaction 53
Figure 3.1. Non-response rates are higher for mental health questions than they are for other variables, and vary substantially across countries 104
Figure 3.2. Young people, men and those with lower levels of education are less likely to respond to mental health questions 106
Figure 3.3. Prevalence of psychological distress and depressive symptoms risk varies by as much as 100 percent across European OECD countries 108
Figure 3.4. Are differences in reported outcomes by gender and race/ethnicity due to differences in underlying mental health or to measurement issues? 109
Figure 3.5. Symptoms of depression rose substantially in eight European OECD countries in the first year of the pandemic 113
Figure 3.6. Until 2019, mental health improved somewhat in European OECD countries, and remained roughly stable in the United States, despite greater mental health awareness 115
Figure 3.7. Shifts from in-person to telephone-administered surveys are associated with deteriorations in negative affect balance 119
Figure 3.8. Positive mental health scales may better approximate a normal distribution 122
Figure 3.9. Different scoring conventions can lead to different estimates of prevalence 124

TABLES

Table 1. ISO codes for countries and world regions 11
Table 2. List of acronyms and abbreviations used in the report 11
Table 1.1. Different mental health conditions influence the extent to which positive mental health is achievable 22
Table 2.1. Advantages and limitations of different tools to measure mental health 41
Table 2.2. Many countries have introduced child and youth surveys, or survey modules, with a mental health focus 48
Table 2.3. Over half of OECD countries reported increasing the collection of mental health data during the COVID-19 pandemic 48
Table 2.4. Overview of structured interviews to monitor mental health conditions 55
Table 2.5. Overview of validated screening tools to monitor both general mental ill-health and risk for specific mental health conditions 55
Table 2.6. Overview of questions about previous diagnoses 58
Table 2.7. Overview of questions about experienced symptoms and mental health conditions 60
Table 2.8. Overview of questions about suicidal ideation and suicide attempts 63
Table 2.9. Overview of questions about general mental health status 64
Table 2.10. Overview of indicators of positive mental health 66
Table 2.11. High-level overview of additional mental health-related topics collected by countries 69
Table 2.12. MHI-5 Questionnaire with scoring breakdown 72
Table 2.13. Kessler Scale 10/6 Questionnaire with scoring breakdown 73
Table 2.14. PHQ-9/8 questionnaire with scoring breakdown 74
Table 2.15. GAD-7/GAD-2 Questionnaire with scoring breakdown 75

Table 2.16. PHQ-4 Questionnaire with scoring breakdown 75
Table 2.17. WG-SS Enhanced Questionnaire 76
Table 2.18. AUDIT/ AUDIT-C Questionnaire with scoring breakdown 77
Table 2.19. OECD core questions on subjective well-being 78
Table 2.20. WHO-5 questionnaire with scoring breakdown 78
Table 2.21. SF-36 vitality subscale questionnaire with scoring breakdown 79
Table 2.22. SWLS questionnaire with scoring breakdown 79
Table 2.23. MHC-SF questionnaire with scoring breakdown 80
Table 2.24. (S)WEMWBS questionnaire with scoring breakdown 81
Table 3.1. Overview of mental health tool performance on statistical quality, data collection and analysis metrics 93
Table 3.2. The relationship between stigma and prevalence is complex, but in some instances, stigma may lead to lower reported prevalence of mental disorders 105
Table 3.3. The correlation of answers to similar questions depends on whether the phrasing is positive or negative 117

Reader's Guide

Table 1. ISO codes for countries and world regions

AUS	Australia	**FIN**	Finland	**LVA**	Latvia
AUT	Austria	**FRA**	France	**MEX**	Mexico
BEL	Belgium	**GBR**	United Kingdom	**NLD**	Netherlands
CAN	Canada	**GRC**	Greece	**NOR**	Norway
CHE	Switzerland	**HUN**	Hungary	**NZL**	New Zealand
CHL	Chile	**IRL**	Ireland	**OECD**	OECD average
COL	Colombia	**ISL**	Iceland	**POL**	Poland
CRI	Costa Rica	**ISR**	Israel	**PRT**	Portugal
CZE	Czech Republic	**ITA**	Italy	**SVK**	Slovak Republic
DEU	Germany	**JPN**	Japan	**SVN**	Slovenia
DNK	Denmark	**KOR**	Korea	**SWE**	Sweden
ESP	Spain	**LTU**	Lithuania	**TUR**	Türkiye
EST	Estonia	**LUX**	Luxembourg	**USA**	United States

Table 2. List of acronyms and abbreviations used in the report

ACASI	Audio computer-assisted self-interviewing
ADHD	Attention deficit hyperactivity disorder
ASD	Autism spectrum disorder
AUDIT	Alcohol Use Disorders Identification Test
AUDIT-C	Alcohol Use Disorders Identification Test – Concise
CAGE	CAGE Substance Abuse Screening Tool
CES-D	Center for Epidemiological Studies Depression Scale
CFA	Confirmatory factor analysis
CIDI	Composite International Diagnostic Interview
CIUS	Compulsive Internet Use Scale
COVID-19	Coronavirus disease of 2019
DASS	Depression, Anxiety and Stress Scale
DAWBA	Development and Wellbeing Assessment – Eating Disorder Module
DAWMA	Development and Well-being Assessment
DSM	Diagnostic and Statistical Manual of Mental Disorders
EPDS	Edinburgh Post-natal Depression Scale
EFA	Exploratory factor analysis
EHIS	European Health Interview Survey
EU-SILC	European Union Statistics on Income and Living Conditions Survey
EVI	Energy and Vitality Index
GAD	Generalized Anxiety Disorder
GHQ	General Health Questionnaire
HADS	Hopkins Symptom Checklist

HRQoL	Health-related quality of life
IES-R	Impact of Event Scale – Revised
I-PAPI	Interviewer-administered paper and pencil surveys
K10 (K6)	Kessler Scale
MDE	Major depressive disorder
MHC	Mental health component summary
MHC-LF	Mental Health Continuum – Long Form
MHC-SF	Mental Health Continuum – Short Form
MHI	Mental Health Inventory
Mini-SPIN	Mini-Social Phobia Inventory
NHS	National Health Service
NSO	National statistical office
OECD	Organisation for Economic Co-operation and Development
PCL	PTSD Checklist
PC-PTSD-5	Primary Care PTSD Screen for DSM-5
PCS	Physical component summary
PHQ	Patient Health Questionnaire
PHQ-ED	Patient Health Questionnaire – Eating Disorder Module
PHQ-PD	Patient Health Questionnaire – Panic Disorder
PSS	Cohen Perceived Stress Scale
PTSD	Post-traumatic stress disorder
SCARED	Screen for Child Anxiety and Related Emotional Disorders
SCOFF	Sick, Control, One, Fat, Food (eating disorder questionnaire)
SDQ	Strengths and Difficulties Questionnaire
SF-12 (-36)	Short Form Health Status
SMFQ	Short Moods and Feelings Questionnaire
SRMH	Self-reported mental health
STAI	The State and Trait Anxiety Scale
(S)WEMWBS	(Short) Warwick-Edinburgh Mental Well-being Scale
SWLS	Satisfaction with Life Scale
ROC	Receiver operating characteristic curve
WG-ES	Washington Group Extended Set on Functioning
WG-SS	Washington Group on Disability Statistics Short Set on Functioning
WHO	World Health Organization
WHO-5	WHO-5 Wellbeing Index
WHOQOL-BREF	WHO Quality of Life – BREF

Executive summary

Mental health is a vital component of people's broader well-being…

Mental health plays a central role in people's lives and is intrinsically tied to many other aspects of people's wider well-being. This was underscored during the COVID-19 pandemic, when direct health impacts and loss of lives combined with social isolation, loss of work and financial insecurity all contributed to a significant worsening of people's mental health. Data from 15 OECD countries suggest that by late 2020, over one-quarter of people experienced symptoms of depression or anxiety. Already, well before the pandemic hit, it was estimated that half of the population will experience a mental health condition at least once in their lifetime and the economic costs of mental ill-health amounted to more than 4% of GDP annually. Good mental health, on the other hand, can boost people's resilience to stress, help them realise their goals and actively contribute to their communities. Positive mental health, or having high levels of emotional and psychological well-being, is also increasingly being recognised as policy target in its own right by health and other government agencies across the OECD.

…but guidance on how to best monitor it at the population level is lacking

It is essential for governments interested in improving mental health to monitor outcomes for both ill-health and positive mental health at the broader population level. Statistics that only consider people diagnosed or treated by health care professionals are strongly affected by how accessible and developed a country's health care system is, and identifying at-risk groups early on requires tracking outcomes well before a person engages with health care services. Moreover, good mental health is a foundational asset for the population, and as such, is valuable to track in its own right. Successful mental health promotion strategies also require understanding of how broader risk and resilience factors, such as people's material conditions, quality of life and social relationships (and inequalities in these), impact their mental health. Data on these topics are typically collected in (social) population survey statistics that can be expanded to include mental health outcomes, to support this greater understanding and provide a better evidence base for policy.

Internationally, data on population-wide mental health outcomes are increasingly available but remain infrequently collected and poorly harmonised across countries. Several of the population mental health statistics the OECD is regularly publishing in its long-standing effort to promote a society-wide response to improving mental health are only available on a regular basis for a subset of OECD countries, are more than five years old at the time of publication for several countries, and in some cases stem from non-official data sources.

How is this report intended to be used?

This report aims to support national statistical offices and other data producers in collecting high-quality measures of population mental health outcomes in a more frequent, consistent and internationally harmonised manner. The OECD took stock of what member countries are already doing in this area with

a questionnaire that was shared with the OECD Committee on Statistics and Statistical Policy in February 2022. Almost all national statistical offices, and in many cases also health agencies, reported back. The report documents existing measurement practice to identify where countries are converging when it comes to gathering population mental health outcomes, and where gaps remain. In addition, available measurement tools are assessed to provide recommendations for priority measures official data producers can adopt in household, health and social surveys.

Key messages and recommendations

- **Collecting data on both mental ill-health and positive mental health at the population level** would yield a more complete picture of mental health. Integrating relevant questions in population surveys that include information on other aspects of people's lives would help better understand the drivers and policy levers associated with improving mental health outcomes. This can provide new avenues for proactive rather than reactive policy design, and mental health strategies that both reduce ill-health and promote good emotional and psychological well-being.
- **The pandemic has spurred new efforts in mental health data collection, and it will be important to keep up the momentum.** Since March 2020, 7 in 10 OECD countries added mental health modules to existing surveys or launched new mental health surveys, many of them administered multiple times per year, or even more frequently. It is unclear whether these will continue in the future. A return to business as usual prior to the pandemic would mean that half of OECD countries only collect mental health data every four to ten years.
- **Some aspects of mental health are more frequently captured than others, and there is scope for better cross-country harmonisation of measures.** This is in particular reflected by lack of harmonisation when it comes to measuring symptoms of anxiety, affect and eudaimonia (i.e. a sense of meaning and purpose in life), and very uneven use of tools that assess specific mental health conditions beyond anxiety and depressive disorders.
- **The report suggests adding four specific tools in relevant population surveys to build a small set of more internationally harmonised population mental health indicators.** These recommendations have been formed based on a comparative assessment of their statistical quality, their response burden and cost, and existing data collection practices. They do not imply the phasing out of other tools that OECD countries are already using to capture population mental health outcomes.
    - *Mental ill-health – priority recommendation*: The Patient Health Questionnaire-4 (PHQ-4) could be included in more frequent surveys, alongside the regular collection of the PHQ-8 or PHQ-9 in health surveys. It covers symptoms of both depression and anxiety, and does so with only four questions.
    - *Positive mental health – recommendation*: Based on trends in country measurement practice, either the WHO-5 or SWEMWBS could be used to measure affective and eudaimonic aspects of positive mental health in a comparative way. The topic of measuring affect and eudaimonia specifically will continue to be explored in future OECD work on subjective well-being.
    - *General mental health status – recommendation*: Similar to commonly used questions that ask respondents to rate their physical health, a single question about a respondent's general mental health status could be included in a range of surveys across a country's broader data infrastructure system. Over half of countries include such questions already, though question wording varies widely. The following framing has been adopted by at least three OECD countries: "In general, how is your mental health? Excellent / Very good / Good / Fair / Poor."

1 What is population mental health and why should we measure it?

Mental health is a vital component of people's well-being, and measuring it is essential to monitor what ultimately matters to people. The aim of this report is to encourage official data producers to collect data on population mental health outcomes more frequently and in an internationally harmonised manner. Considering all aspects of mental health, ranging from mental ill-health (which may or may not include a diagnosed mental health condition) to positive mental states, can provide new avenues for a proactive rather than reactive design of mental health systems and services, and it can open up space for policy to focus on both reducing illness and promoting people's flourishing. Collecting data on mental ill-health and positive mental health in household, social and health surveys would yield a more complete picture of mental health and help to better understand the drivers and policy levers for improving it.

Good mental health is a vital component of people's well-being. Good mental health enables individuals to realise their own potential, cope with the stresses of life, work productively and make a positive contribution to their communities (World Health Organization, 2019[1]). Mental ill-health on the other hand accounts for one of the largest and fastest-growing categories of the burden of disease worldwide; its economic costs, including investment in the mental health system and lower employment and productivity, are estimated at more than 4% of GDP in OECD countries (Rehm and Shield, 2019[2]; OECD, 2021[3]). In 2016, already well before the COVID-19 pandemic, deaths of despair (due to suicide, acute alcohol abuse or drug overdose) were one of the largest causes of preventable deaths in OECD countries, six times higher on average than deaths due to homicide, and three times higher than road deaths (OECD, 2020[4]). Over the period 2016-18, one in eight people living in OECD countries experienced more negative than positive emotions during a typical day (OECD, 2020[4]).

Mental health has come to the forefront of the public debate during COVID-19. Besides the direct effect of the pandemic in terms of the high number of lives lost, social isolation, loss of work and financial insecurity all led to a significant worsening of people's mental health, with more than a quarter of people in 15 OECD countries experiencing symptoms of anxiety or depression by late 2020 (OECD, 2021[3]; OECD, 2021[5]). Populations living in vulnerable situations, including women, young people, those in precarious employment and financial situations, racial and ethnic minorities, and people living with existing mental health conditions and substance use disorders, have been particularly affected.

While it is clear that mental health matters for people's well-being, and that substantial parts of the population are living with and affected by mental ill-health, discussion so far have not focused sufficiently on how governments should best monitor it at the broader population level, and on how to consider both mental ill-health and positive mental states. This also requires a conversation about what exactly is meant by "mental health" and about which outcomes are most relevant for policy makers responsible for treatment, prevention and promotion strategies.

This chapter first makes the case for why regular measures of population-level mental health outcomes should be collected. It then presents how different components of mental health, including mental ill-health and positive mental health, have been distinguished in research and practice.[1] This provides the basis for a common understanding and terminology used throughout this report, including in the subsequent chapters on available measurement tools and current measurement activities in OECD countries (Chapter 2) and on what is known about their statistical quality and measurement practice (Chapter 3).

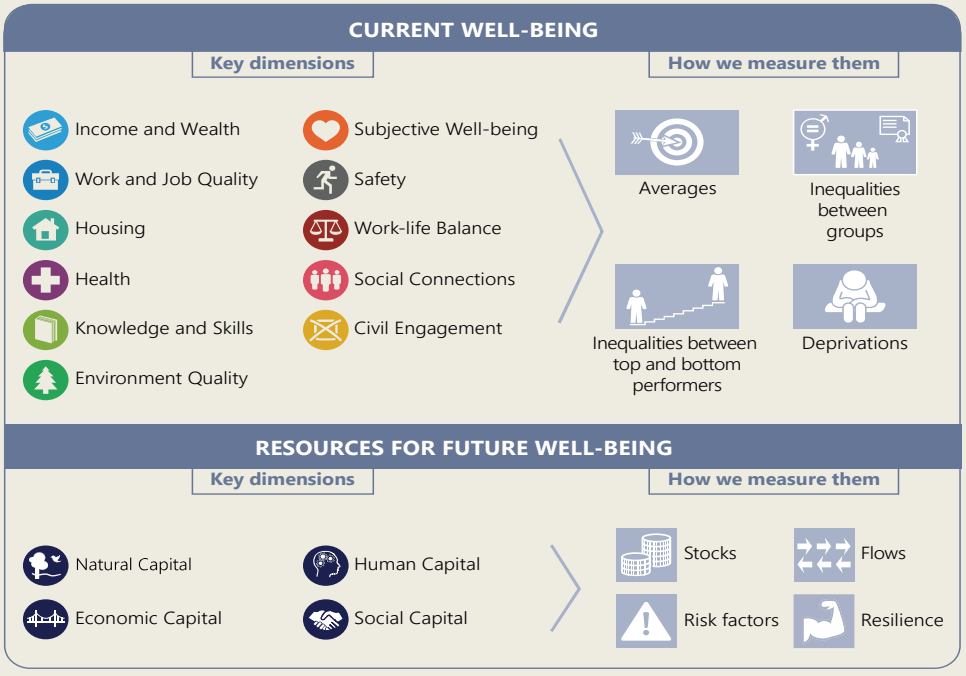
The importance of focusing on population mental health outcomes

The OECD has a long record of collating international health statistics and promoting a society-wide response to improving mental health. This includes the *2015 OECD Recommendation on Integrated Mental Health, Skills and Work Policy* and its follow-up report, *Fitter Minds, Fitter Jobs*, as well as the recent publication *A New Benchmark for Mental Health Systems,* which sets out a framework for understanding mental health performance and assesses whether countries are delivering the policies and services that matter for health system performance (OECD, 2021[3]; OECD, 2015[6]; OECD, 2021[7]). Preventing mental illness, promoting mental well-being and taking a multisectoral approach to mental health are amongst the key principles of the OECD's *New Benchmark* framework, and a number of population-level outcomes indicators are included under these principles (life satisfaction, suicide rate and inequalities in mental distress by education and employment status). In addition, the OECD *How's Life?* reports (which assess well-being, inequality and sustainability in over 40 member and partner countries, see Box 1.1) also feature a range of outcome indicators relevant to mental health. However, several of these are produced irregularly, only cover a subset of OECD countries and in some cases are drawn from non-official sources.[2]

Box 1.1. Measuring people's well-being

The OECD Well-being Framework is a broad outcome-focused tool to measure human and societal conditions and assess whether life as a whole is getting better for people living in OECD countries (OECD, 2020[4]). It includes both current well-being in the "here and now", which focuses on living conditions at the individual, household and community levels, and systemic resources needed to sustain well-being in the future.[3] The Well-being Framework underpins the OECD *How's Life?* report series and a wide range of other OECD work related to well-being (for an overview, see https://www.oecd.org/wise/).

Figure 1.1. The OECD Well-being Framework



Source: OECD (2020[4]), *How's Life? 2020: Measuring Well-being*, OECD Publishing, Paris, https://doi.org/10.1787/23089679.

Mental health is not explicitly identified as a separate dimension of well-being in the framework, but mental health outcomes are relevant to several dimensions:

- First, and foremost, the broad "health" dimension of the Framework encompasses both mental and physical health. For example, two indicators of mental ill-health (deaths from suicide, acute alcohol abuse and drug overdose, and the share of people at risk of depression) were included in the OECD *How's Life? 2020* report under this dimension.
- Second, the "subjective well-being" dimension encompasses elements of good psychological functioning, notably eudaimonia and positive and negative affect. People's own evaluation of their lives (e.g. life satisfaction) is also included here.
- Last, "Human capital", included under resources for future well-being, refers to "the knowledge, competencies, skills and health status of individuals, which are viewed here from the perspective of their contribution to future well-being" and includes indicators such as premature mortality and obesity prevalence (OECD, 2020[4]).[4]

In addition, several aspects of positive functioning that are often included in broad definitions of positive mental health, such as social connections, financial security (income and wealth), and knowledge and skills, are captured by separate dimensions within the OECD Well-being Framework.

The aim of this report is to encourage official data producers to collect data on population-level mental health outcomes more frequently and in an internationally harmonised manner. This is in line with a well-being approach that assesses what ultimately matters to people themselves and their capabilities to live a life of their choosing (in this case, feeling mentally healthy and free of mental distress) (OECD, 2020[4]). Moreover, several well-being drivers measured by more frequently available input or output indicators may be imperfectly correlated with such outcomes (e.g. mental health expenditure is a poor proxy of mental health status if the health care system is inaccessible; similarly, the number of drugs prescribed says little about people's (mental) health conditions) (OECD, 2011[8]).

Collecting data on mental health status for the entire population, rather than only for people diagnosed or treated by health care professionals, is important for a number of reasons. First, measures focusing on the numbers diagnosed might only reflect how accessible and developed a country's health care system is, and how likely people (and certain population groups) are to seek treatment. Second, strategies to prevent mental ill-health would benefit from identifying at-risk groups early on. So, they necessitate tracking outcomes prior to, and following, engagement with the health system. Third, positive mental health is a foundational asset for the population, and as such, is valuable to track in its own right Linking mental health with the broader risk and resilience factors typically also collected in population (survey) statistics, such as people's material conditions, quality of life and social relationships (and inequalities in these), can equally support mental health strategies.

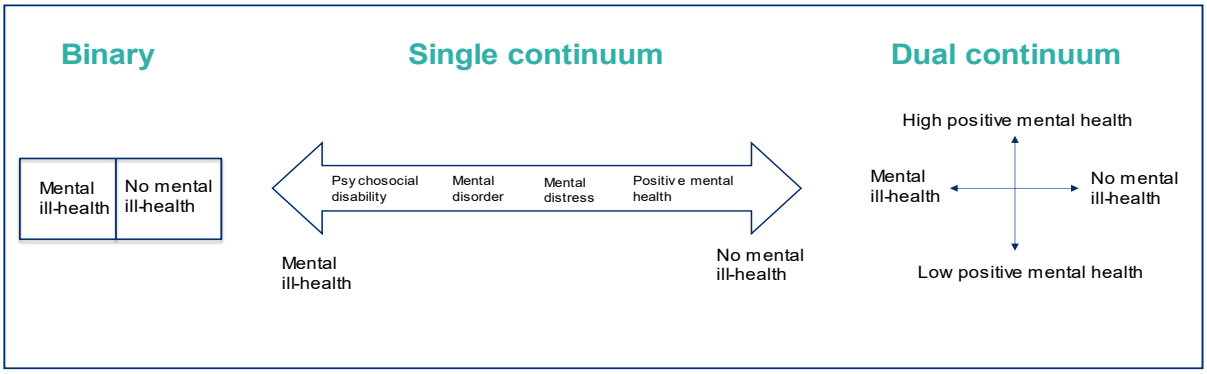
Concepts of mental health: From illness to wellness

Previous OECD work on mental health has adopted the widely accepted definition of mental health by the World Health Organisation (WHO): "a state of well-being in which the individual realises his or her abilities, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to his or her community" (OECD, 2021[3]; OECD, 2015[6]). This definition explicitly states that mental health is not the absence of illness and encompasses multiple aspects of psychosocial functioning (World Health Organization, 1948[9]).[5]

Various theories about what mental health entails have been developed over the past decades. These range from those focusing on symptoms of mental illness either being present or not ("binary model"), to those conceiving of mental health as a spectrum of experience ("single-continuum model"), all the way to viewing mental ill-health and positive mental health as related but distinct experiences ("dual-continuum model") (Figure 1.2). Each of these models carries different implications for which mental health outcomes need to be tracked in order in order to capture the concept in its entirety.

Figure 1.2. Models of mental health

Stylised conceptual frameworks of mental health



Source: Adapted from Iasiello et al. (2020[10]), "Mental Health and/or Mental Illness: A Scoping Review of the Evidence and Implications of the Dual-Continua Model of Mental Health", Evidence Base, 10.21307/eb-2020-001.; Keyes, C. (2005[11]), "Mental illness and/or mental health? Investigating axioms of the complete state model of health", Journal of Consulting and Clinical Psychology, 73(3): 539.

The binary model

Clinical psychology, psychiatry and research more generally have historically focused on the reduction of mental illness symptoms, or psychopathology, in order to improve mental health. In this "disease-centred" perspective, mental illness (in the form of conditions defined by psychiatric classification systems) is the focal concept, and the goal of intervention is primarily to help reduce the associated symptomatology, rather than to support people into wellness. In this perspective, an individual is capable of experiencing one of two alternative states: either being diagnosed as mentally ill or being presumed mentally healthy (Routledge et al., 2016[12]; Keyes, 2005[11]; Trent, 1992[13]).

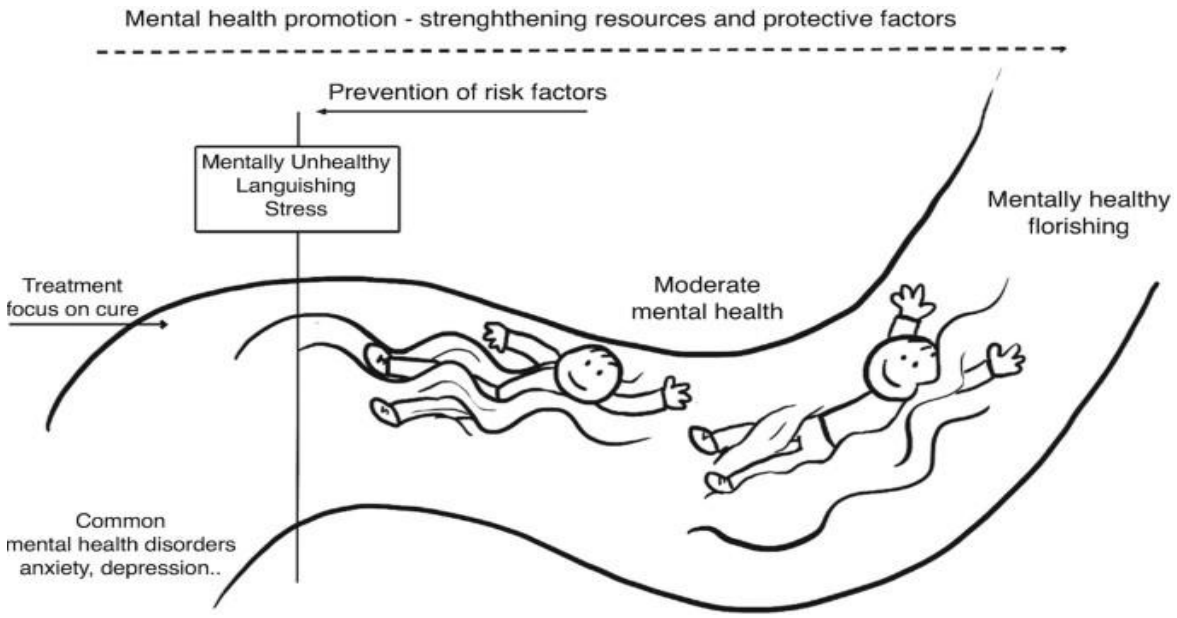
Binary categorisations of mental illness can be useful, for instance, when a person is trying to access appropriate health care or other support services or for defining guidelines and treatment pathways to manage diagnosed conditions. However, practitioners and researchers have criticised the reductionist nature of this model, i.e. the notion of an arbitrary point where illness transitions to full health and the presumed impossibility of "gaining" more mental health once the threshold of no diagnosable condition is crossed (Herron and Trent, 2000[14]).

Mental health as a continuum

An alternative approach is to characterise mental health as a continuum of experience, from severe mental ill-health, on one end of the spectrum, through to positive mental health (high levels of emotional and psychological well-being) on the other (Patel et al., 2018[15]; Payton, 2009[16]; Greenblat, 2000[17]). This view is rooted in a "salutogenic" approach that focuses on factors that support health and well-being, beyond the traditional focus on risks, symptoms and problems. It acknowledges a wider breadth of people's experiences (which are different for someone who might feel worried or has trouble sleeping compared to a person experiencing a full-blown episode of major depression).

In this model, "everyone has mental health", and an individual can move up and down the spectrum throughout their life (including up to a daily or weekly basis) depending on the context they find themselves in, the challenges they face and the internal and external resources available to them. Some researchers have used the metaphor of a river, rather than a linear continuum, to express this constant process and the fluidity of different states between acute mental ill-health and positive mental health (Figure 1.3 (Koushede and Donovan, 2022[18])

Figure 1.3. The mental health continuum as a river



Source: Koushede, V. and R. Donovan (2022[18]), "Applying Salutogenesis in Community-Wide Mental Health Promotion", *The Handbook of Salutogenesis*. Springer, Cham. https://doi.org/10.1007/978-3-030-79515-3_44.

From a policy perspective, considering the full spectrum between mental ill-health and positive mental health carries implications for both targeting and designing interventions and can provide new avenues for proactive rather than reactive system and service design. The single-continuum model adds value vis-à-vis the binary perspective by providing the space for mental health strategies to focus not just on "curing" (diagnosed) illness or reducing the associated symptoms, but also on preventing people in the middle of the spectrum from doing worse and on promoting mentally healthy populations.

Mental ill-health and positive mental health as a dual continuum

A third conceptual view, increasingly considered by international players such as the World Health Organisation, several public health agencies, national statistical offices and other government departments, more clearly differentiates between mental ill-health, on one side, and positive mental health, on the other (Statistics Canada, n.d.[19]; Australian Early Development Census, 2012[20]; Swiss Health Observatory, n.d.[21]; Government of Western Australia Mental Health Commission, 2021[22]; Queensland Government, 2015[23]) (World Health Organization, 2022[24]). This "dual-continuum" model characterises mental ill-health and positive mental states as related but distinct experiences (placing them on two different but intersecting continua), rather than as extreme ends of a single spectrum.[6]

Mental ill-health and positive mental health, or high levels of emotional and psychological well-being, are closely interconnected. Gains in good mental health at the population level imply declines in average mental disorders over time, while experiencing positive mental health decreases the risk of developing a mental disorder, can help recovery once it has been developed and is thus considered an important resilience factor (Keyes, Dhingra and Simoes, 2010[25]; Robinson, 2012[26]; Santini et al., 2022[27]).

Proponents of the dual-continuum model, however, argue that the association between ill-health and positive mental health is not linear, as the single-continuum model might suggest: the mere absence of *clinically significant* symptoms of mental ill-health, or diagnosed conditions, does not always imply a thriving mental state. Conversely, a person could have symptoms of a mental disorder and associated

distress and disability, but also be satisfied with their life as a whole and achieving their potential (Galderisi et al., 2015[28]). This view, which aims to acknowledge the full diversity of human experiences, is also often echoed by people with lived experience of mental health conditions (New Zealand Initial Mental Health and Wellbeing Commission, 2020[29]).

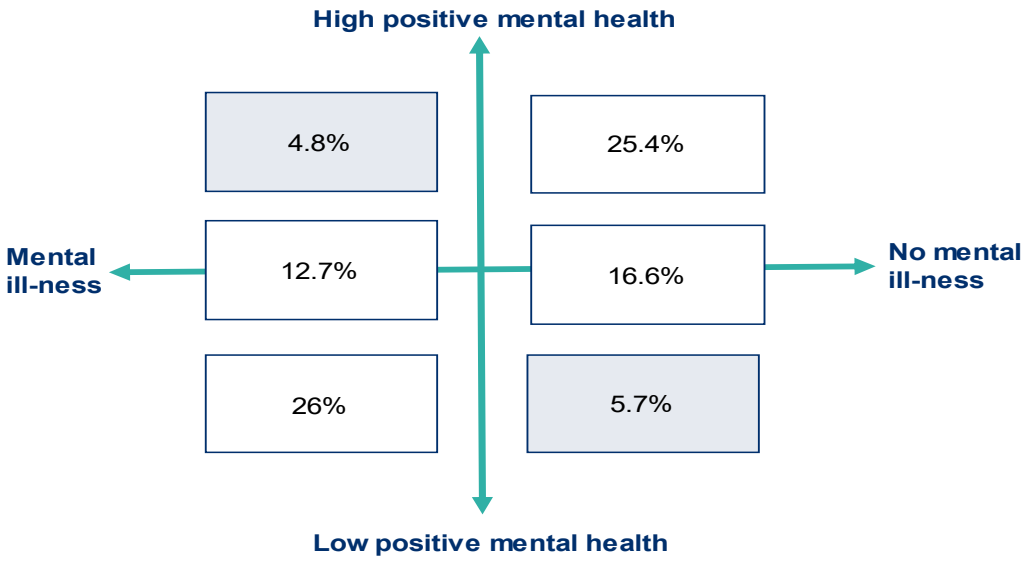
The majority of research supporting a dual continuum has relied on confirmatory factor analysis (CFA) to compare whether survey data best fit a single- or dual-continuum model. Keyes (2005[11])measured aspects of emotional, psychological and social well-being[7] and some common forms of mental illness (presence of a major depressive episode, generalised anxiety disorder, panic disorder, or alcohol dependence in the past year) in a nationally representative sample of US adults. He then used CFA to highlight the existence of two correlated but separate latent factors. Additional studies of non-US populations using a variety of measurement tools for both positive mental health and mental illness have further supported the notion of the dual-continuum model. A recent review identified 83 peer-reviewed empirical articles, including cross-sectional, longitudinal and intervention studies, which provided support for the superior explanatory power of dual-continuum models of mental health over a single-continuum model (Iasiello, van Agteren and Cochrane, 2020[10]; Routledge et al., 2016[12]).[8]

The typical visualisation of two completely *orthogonal* axes in the dual-continuum model can, however, be misleading. Several studies classify individuals into separate groups around the model's quadrants, using variations of categories such as "complete mental health" (no mental illness, high positive mental health), "vulnerable" (low mental illness, low positive mental health), "symptomatic but content" (high mental illness, high positive mental health) and "struggling" (high mental illness, low positive mental health) (Iasiello, van Agteren and Cochrane, 2020[10]). Distributions within these categories, however, strongly suggest that levels of positive mental health and mental ill-health are highly related and that mental health conditions bring significant impairments for emotional and psychological well-being. For instance, a study of Australian schoolchildren shows that only around 5% of children experience either high levels of positive mental health but also mental ill-health, or low levels of positive mental health but no mental ill-health (Figure 1.4). Similarly, while in a study by Keyes only one in five people who had *no* diagnosed mental health condition in the past year recorded high positive mental health, even fewer respondents *with* a mental disorder were likely to do so (Table 1.1) (Keyes, 2005[11]).

In the same study, experiences of positive mental health also vary strongly according to the type of psychological disorder experienced in the past year (and its severity at the time of the survey), ranging from only 2% for those with generalised anxiety disorder to 8% for those who were alcohol-dependent (Table 1.1). Nevertheless, the share of respondents with a high degree of mental ill-health who can attain some degree of positive mental health is not insignificant. Lesser-known interlinkages between various aspects of emotional and psychological well-being and different, even severe, mental health conditions are also possible: some studies suggest that, compared to psychologically-healthy adults, people with depression might react to negative events with less distress, while people with bipolar disorder experience greater positive emotions during mania, people with schizophrenia can construct meaning from their hallucinations and delusions, and trauma survivors can live meaningful lives upon coping with their stressful experiences (Goodman, Doorley and Kashdan, 2018[30]).

Figure 1.4. The dual continuum of mental health in Australian children

Share of children in their first year of formal full-time schooling experiencing different degrees of mental ill-health (anxiety, depression, behaviour problems) and positive mental health (psychosocial functioning), Australia, 2012



Note: Data are drawn from the 2012 national Australian Early Development Index, and responses about children were provided by their school teachers. Darker shaded fields refer the share of children who have either low positive mental health but no mental illness or those who experience mental illness but also high positive mental health. The original source termed these categories as mental health difficulties (e.g. anxiety disorders, depression, behavioural problems) and mental health competency (e.g. healthy psychosocial functioning).
Source: Australian Early Development Census (2012[20]), *The mental health of Australian children: A dual continuum*, https://www.aedc.gov.au/resources/detail/the-mental-health-of-australian-children-a-dual-continuum.

Table 1.1. Different mental health conditions influence the extent to which positive mental health is achievable

Share of adults with a mental health condition in the past 12 months that report low, moderate or high positive mental health, United States, 1994-95

	Low positive mental health (languishing)	Moderate positive mental health	High positive mental health (flourishing)
Overall sample (n=3032)	16.9%	65.1%	18%
Major depressive episode (n=422)	33.9%	60.2%	6.2%
Generalised anxiety disorder (n=98)	55.1%	41.8%	2%
Panic disorder (n=204)	39.2%	58.3%	2.5%
Alcohol dependence (n=194)	24.7%	69.1%	7.7%
Comorbidity (n=193)	43.5%	55.4%	2.1%

Note: Data are drawn from the "Midlife in the United States" study. Mental disorders were measured by the Composite International Diagnostic Interview Short Form (CIDI-SF) scale. Flourishing (languishing) was defined as an individual exhibiting high (low) levels on one of two questions about positive affect and high (low) levels on six of 11 questions about positive functioning (per Ryff's scales of psychological well-being and Keyes' scales of social well-being). All other respondents were categorised within moderate positive mental health. Comorbidity refers to the experience of more than one mental health condition, regardless of in which combination.
Source: Keyes, C. (2005[11]), "Mental illness and/or mental health? Investigating axioms of the complete state model of health", *Journal of Consulting and Clinical Psychology*, 73(3): 539.

By distinguishing between mental ill-health and positive mental health, the dual-continuum model also implicitly suggests that the relative importance of their respective drivers differs. This is important for policy and clinical practice, as the same strategies for preventing mental illness might not be sufficient for enhancing positive mental states, and vice versa. Evidence on this is still emerging. Some population-based studies from Denmark and England have suggested that deprivations in people's material conditions and quality of life (such as low income and educational attainment, lack of employment and financial insecurity) predict outcomes at the tail-end of each continuum (i.e. both mental ill-health and low levels of positive mental health). These same socio-economic factors did not play an equally strong role in determining high levels of positive mental health (Stewart-Brown et al., 2015[31]; Nielsen et al., 2016[32]; Santini et al., 2020[33]). However, population-based data from Canada and Slovenia suggests that higher financial security and household income are indeed associated with increased odds of psychological well-being (Varin et al., 2020[34]; Vinko et al., 2022[35]). By contrast, relational factors such as greater social connectedness, improved family relations and participation in recreational activities have been associated with both reduced risk of mental health conditions as well as higher positive mental health in the majority of studies (Van Lente et al., 2012[36]; Santini et al., 2020[33]; Santini et al., 2017[37]; Solin et al., 2019[38]; Thoits, 2011[39]).

The value-add of the dual-continuum model (over the single continuum) is that it more explicitly communicates that both mental ill-health and good mental states impact people's lives. From a measurement perspective, the dual-continuum model suggests that collecting data on both mental health and positive functioning in population surveys and health assessments would yield a more complete picture of mental health. This would also help to identify the factors, and by extension policy levers, associated with the dual goals of improving positive mental health and reducing mental illness. This report hence considers the two constructs separately where possible and defines each in more detail in the following sections.

Mental ill-health

The term mental ill-health refers to diagnosable mental and behavioural conditions, as well as the transdiagnostic characteristic of psychological distress.

Mental health conditions

The terms "conditions" or "disorders" are used in this report to describe symptoms reaching the clinical threshold of a diagnosis according to psychiatric classification systems such as the World Health Organization International Classification of Disease (ICD) or the American Psychiatric Association Diagnostic and Statistical Manual (DSM).[9] There are more than one hundred separate diagnoses and disorders featured in these classification systems, including mild or moderate anxiety and depression, drug and alcohol use disorders, and severe disorders such as severe depression, bipolar disorders and schizophrenia, each with their own specific symptoms, age of onset and trajectory (Box 1.2). The experience of mental health conditions can be highly fluid both over the life-course and over much shorter periods of several weeks – e.g. an individual experiencing a moderate depressive episode can worsen so

that the condition becomes severe, just as a severe episode can be stabilised with the symptoms lessened or alleviated (OECD, 2021[3]).

It is estimated that half of the population will experience a mental health condition in their lifetime and about one in five people in any given year (OECD, 2019[40]). The data currently available from population-based surveys often focus on experiencing symptoms of anxiety and depression (see Chapter 2). Pre-COVID-19 point estimates from the Institute for Health Metrics and Evaluation (IHME) suggest that the most common mental disorder in EU countries is anxiety disorder, with an estimated 25 million people (or 5.4% of the population) living with this condition in 2016, followed by depressive disorders, which affected over 21 million people (or 4.5% of the population). An estimated 11 million people across EU countries (2.4%) have drug and alcohol use disorders. Severe mental illnesses such as bipolar disorders affected almost 5 million people (1% of the population), while schizophrenic disorders affected 1.5 million people (0.3%) (OECD/European Union, 2018[41]).[10]

Box 1.2. Examples of mental health conditions and their symptoms

According to the DSM, a mental health condition is a syndrome characterised by a clinically significant disturbance in an individual's cognition, emotion regulation or behaviour that reflects a dysfunction in the psychological, biological or development processes underlying mental functioning. Mental disorders are usually associated with significant distress or disability in social, occupational or other important activities (American Psychiatric Association, 2013[42]). Comorbidity of mental disorders and physical illnesses and multiple mental health problems is common. The most recent version, DSM-5, was published in 2013 and lists a total of 157 diagnoses and close to 300 disorders. Some of the most common clusters of disorders featured include:

Mood/affective disorders

Mood disorders, or affective disorders, are characterised by a disturbance of the general emotional state that interferes with an individual's ability to function. Various forms of mood disorders exist: for instance, a major depressive disorder is characterised by persistent periods of low mood, low self-esteem and loss of interest in usually pleasurable activities lasting at least two weeks. Physical symptoms such as fatigue, headaches or digestive problems are also common. Bipolar disorder is characterised by alternating periods of depression and periods of mania (pathologically elevated mood, arousal and energy levels).

Anxiety disorders

Anxiety disorders are characterised by excessive and uncontrollable feelings of anxiety and fear. Specific symptoms depend on the type of anxiety disorder present. The most common anxiety disorders are generalised anxiety disorder, panic disorder and social anxiety disorders. In addition, various specific phobias (a fear of specific objects or situations) exist, like intense fear of heights or of flying.

Substance use disorders

Substance use disorder is a condition characterised by an uncontrollable intake of substances despite adverse consequences, and it is often accompanied by emotional, physical and behavioural problems and an inability to stop consuming despite several attempts. For instance, alcohol use disorder is a type of substance abuse disorder and includes frequent and heavy alcohol use.

Adjustment disorders

An adjustment disorder is characterised by a maladaptive emotional or behavioural reaction to a psychosocial stressor. Adjustment disorders occur when individuals have significant difficulties to adjust

or cope with a stressful life event. For example, post-traumatic stress disorder (PTSD) usually develops due to exposure to traumatic life events or threatening situations, such as war, sexual assault or child abuse. Symptoms can range from sleeping difficulties, difficulty concentrating or irritability to hypervigilance and an exaggerated startle response.

Psychotic disorders

Psychotic disorders are severe mental health conditions with delusions and hallucinations as common symptoms. The most common psychotic disorder is schizophrenia, in which people interpret and experience reality abnormally and which is characterised by a combination of hallucinations, delusions and extremely disordered thinking and behaviour that impairs daily functioning.

Personality disorders

Personality disorders are characterised by long-term maladaptive patterns of behaviour, cognition and inner experience that differ significantly from the cultural-social norm. They are associated with difficulties in cognition, emotiveness, interpersonal functioning or impulse control. Three clusters of personality disorders exist: odd or eccentric disorders; dramatic, emotional or erratic disorders; and anxious or fearful disorders.

Somatoform and dissociative disorders

Somatoform disorders are disorders causing physical symptoms that might not be traceable to a somatic cause. Dissociative disorders include problems with memory, awareness, perception or identity; people experiencing dissociative disorder might feel disconnected from their body or develop different identities.

Eating disorders

An eating disorder is characterised by abnormal eating behaviours that affect physical and/or mental health. Various types of eating disorders exist, the most common being bulimia nervosa, anorexia nervosa and binge eating disorder. Eating disorders are often comorbid with anxiety disorders, depression and substance abuse.

Obsessive-compulsive disorder

Obsessive-compulsive disorder is a mental and behavioural disorder characterised by intrusive, reoccurring thoughts or mental images (obsessions) that generate feelings of anxiety, disgust or discomfort, which in turn elicit an urge to perform a certain task or routine, such as hand washing, counting, cleaning or arranging things, in order to relieve this discomfort (compulsions).

Source: (World Health Organization, 2021[43]; American Psychiatric Association, 2013[42]).

Psychological distress

The term “psychological distress” is used in this report to refer to non-specific symptoms of negative affect (such as sadness, anguish, restlessness), sometimes combined with somatic symptoms (such as inability to sleep or loss of appetite) that do not reach the clinical threshold of a diagnosis within psychiatric classification systems.

There is some debate about whether psychological distress and mental conditions form conceptually distinct phenomena. Some researchers have argued that they are qualitatively distinct: psychological distress should only be considered as part of a pathological psychological process and a marker of a mental health condition if it is persistent and in excess of an “expectable response” to adverse events and

other stressors. However, this is difficult to determine in practice and may depend on an individual's socio-economic and overall life conditions (Horwitz, 2007[44]; Phillips, 2009[45]; Payton, 2009[16]; Roger T. Mulder, 2008[46]; Wakefield et al., 2007[47]). The DSM-5 does not provide any criteria for determining when distress becomes clinically significant; an assessment is usually made based on the degree of *impairment* to functioning produced by the distress, rather than its "appropriateness".

Many of the tools developed to assess psychological distress in individuals, as documented in Chapter 2 of this report, are able to reliably distinguish cases of serious mental health conditions from non-serious cases. This suggests that mental disorder and distress, as a transdiagnostic characteristic of most mental health conditions, are indeed closely related (Barlow and Durand, 2009[48]). Moreover, even if the experience of psychological distress were to be temporary, it can imply significant suffering and hardship of individuals and deserves attention in its own right.

Positive mental health

Positive mental health covers psychological, emotional, and in some cases also social, relational and spiritual well-being (Huppert, 2005[49]; Keyes, 2005[11]; Steger et al., 2006[50]; Reis and Gable, 2003[51]).[11]

The concept of positive mental health is closely related to that of subjective well-being, which refers to "good mental states, including all of the various evaluations, positive and negative, that people make of their lives, and the affective reactions of people to their experiences" (OECD, 2013[52]). In 2013, the OECD published *Guidelines on Measuring Subjective Well-being* that identified three broad aspects of subjective well-being (and proposed measures for data collectors):

- *Life evaluation* – a reflective self-assessment of a person's life as a whole, or some specific aspect of it (e.g. life satisfaction measures; satisfaction with financial situation)
- *Affect* – a person's feelings, emotions or states, typically measured with reference to a particular point in time (e.g. measures about experiences of happiness, worry, pain, tiredness)
- *Eudaimonia* – a sense of meaning and purpose in life, or good psychological functioning (e.g. measures of feeling that the things you do in life are worthwhile).

The strongest overlap between positive mental health and subjective well-being tends to be in the area of affect (where common mental health measures emphasise persistent experiences of certain affective states, such as worry, pain or tiredness) and eudaimonia (where many measures were explicitly developed to capture positive mental health). Additional concepts sometimes featured in measures of positive mental health, such as autonomy, optimism, resilience or environmental mastery, are not explicitly referenced in the OECD definition of subjective well-being provided above (Davydov et al., 2010[53]; Snow, 2019[54]; Peterson and Seligman, 2004[55]; Conversano et al., 2010[56]; Ryff and Keyes, 1995[57]). Although these concepts are sometimes included in some (long-form) measures of eudaimonia and psychological functioning that are discussed in the aforementioned *OECD Guidelines* (see Annex 1 and Module D), appraisal styles such as optimism and other character traits are considered mediating factors that influence a person's affective reactions to life circumstances, rather than final well-being outcomes to strive for (OECD, 2013[52]).

The area of greatest conceptual difference between subjective well-being and positive mental health concerns life evaluation measures, which provide a very broad assessment of a person's life in all its dimensions, rather than assessing only their mental health. Nevertheless, in practical terms, life evaluation measures are often included in research on (positive) mental health, since they are valuable as broad outcome measures that reflect a person's perception of their well-being as a whole.

Chapter 2 reviews current data collection practice in OECD countries for the three aspects of subjective well-being mentioned above, as well as for positive mental health summary scales (mostly stemming from

positive psychology) that cover aspects of emotional, psychological and social well-being. Chapter 3, which discusses the statistical quality of mental health tools, focuses only on the latter, since the *OECD Guidelines* have already considered in-depth the issue measuring life evaluations, affect and eudaimonia (OECD, 2013[52]). The topic of measuring affect and eudaimonia specifically will also continue to be explored in future OECD workstreams on subjective well-being. Extremely broad definitions of positive mental health that include domains such as physical and sexual health, financial security, or academic and occupational performance (which are covered elsewhere in the OECD Well-being Framework) are not considered in this publication (Fusar-Poli and Santini, 2022[58]; Fusar-Poli et al., 2020[59]; Harvard Center for Health and Happiness, n.d.[60]).

Conclusion

Measuring mental health is important to fully assess the well-being outcomes that matter to people's lives. The aim of this report is to encourage official data producers to collect population-level data on mental health status more frequently and in an internationally harmonised manner, in order to understand how all societal groups, rather than only those in touch with the health care system, are faring, and to address a topic that is increasingly recognised as public policy challenge.

Mental health is a multifaceted concept that extends beyond a binary distinction between mental illness either being present or not. Considering all aspects of mental health can provide new avenues for the proactive rather than reactive design of mental health systems and services, draw attention to the importance of caring about positive mental health in its own right, and open up the space for policy to focus on both reducing illness and promoting good mental states. Collecting data on both aspects in household, social and health surveys would yield a more complete picture of mental health and help to better understand the drivers and policy levers needed for improving it.

References

American Psychiatric Association (2022), *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (DSM-5-TR™)*, https://www.appi.org/dsm (accessed on 6 April 2022). [64]

American Psychiatric Association (2013), *Diagnostic statistical manual of mental disorder (5th ed.)*. [42]

Australian Early Development Census (2012), *The mental health of Australian children: a dual continuum*, https://www.aedc.gov.au/resources/detail/the-mental-health-of-australian-children-a-dual-continuum (accessed on 17 May 2022). [20]

Barlow, D. and V. Durand (2009), *Abnormal Psychology: An Integrative Approach*, Cengage Learning, https://books.google.com/books/about/Abnormal_Psychology_An_Integrated_Approa.html?id=Mo_q4zFVNo4C (accessed on 7 February 2022). [48]

Conversano, C. et al. (2010), "Optimism and Its Impact on Mental and Physical Well-Being", *Clinical Practice & Epidemiology in Mental Health*, Vol. 6/1, pp. 25-29, https://doi.org/10.2174/1745017901006010025. [56]

Davydov, D. et al. (2010), "Resilience and mental health", *Clinical Psychology Review*, Vol. 30/5, pp. 479-495, https://doi.org/10.1016/j.cpr.2010.03.003. [53]

Exton, C. and L. Fleischer (2023), *The future of the How's Life? dashboard*, OECD Publishing. [61]

Finnish Institute for Health and Welfare (n.d.), *Positive mental health*, https://thl.fi/en/web/mental-health/mental-health-promotion/positive-mental-health (accessed on 7 April 2022). [63]

Fusar-Poli, P. et al. (2020), "What is good mental health? A scoping review", *European Neuropsychopharmacology*, Vol. 31, pp. 33-46, https://doi.org/10.1016/j.euroneuro.2019.12.105. [59]

Fusar-Poli, P. and Z. Santini (2022), "Promoting good mental health in the whole population: The new frontier", *European Neuropsychopharmacology*, Vol. 55, pp. 8-10, https://doi.org/10.1016/j.euroneuro.2021.09.010. [58]

Galderisi, S. et al. (2015), "Toward a new definition of mental health", *World Psychiatry*, Vol. 14/2, pp. 231-233, https://doi.org/10.1002/wps.20231. [28]

Goodman, F., J. Doorley and T. Kashdan (2018), "Well-being and Psychopathology: A Deep Exploration into Positive Emotions, Meaning and Purpose in Life, and Social Relationships", in Diener, E., S. Oishi and L. Tay (eds.), *Handbook of well-being*, DEF Publishers, Salt Lake City. [30]

Government of Canada (n.d.), *Positive Mental Health Surveillance Indicator Framework (PMHSIF)*, https://health-infobase.canada.ca/positive-mental-health/Index (accessed on 8 September 2022). [62]

Government of Western Australia Mental Health Commission (2021), *Western Australian Mental Wellbeing Framework*, https://www.mhc.wa.gov.au/mental-wellbeing-framework/ (accessed on 14 March 2022). [22]

Greenblat, L. (2000), "Understanding Health as a Continuum", *Schizophrenia Bulletin*, Vol. 26/1, pp. 243-245, https://doi.org/10.1093/oxfordjournals.schbul.a033444. [17]

Harvard Center for Health and Happiness (n.d.), *Repository of Positive Psychological Well-Being Scales*, https://www.hsph.harvard.edu/health-happiness/repository-of-positive-psychological-well-being-scales/ (accessed on 7 April 2022). [60]

Herron, S. and D. Trent (2000), "Mental Health: A Secondary Concept to Mental Illness", *Journal of Public Mental Health*, Vol. 2/2, pp. 29-38, https://doi.org/10.1108/17465729200000014. [14]

Horwitz, A. (2007), "Distinguishing distress from disorder as psychological outcomes of stressful social arrangements", *Health: An Interdisciplinary Journal for the Social Study of Health, Illness and Medicine*, Vol. 11/3, pp. 273-289, https://doi.org/10.1177/1363459307077541. [44]

Huppert, F. (2005), "Positive mental health in individuals and populations", in *The Science of Well-Being*, Oxford University Press, https://doi.org/10.1093/acprof:oso/9780198567523.003.0012. [49]

Iasiello, M., J. van Agteren and E. Cochrane (2020), "Mental Health and/or Mental Illness: A Scoping Review of the Evidence and Implications of the Dual-Continua Model of Mental Health", *Evidence Base*, Vol. 2020/1, pp. 1-45, https://doi.org/10.21307/eb-2020-001. [10]

Keyes, C. (2005), "Mental Illness and/or Mental Health? Investigating Axioms of the Complete State Model of Health.", *Journal of Consulting and Clinical Psychology*, Vol. 73/3, pp. 539-548, https://doi.org/10.1037/0022-006X.73.3.539. [11]

Keyes, C., S. Dhingra and E. Simoes (2010), "Change in Level of Positive Mental Health as a Predictor of Future Risk of Mental Illness", *American Journal of Public Health*, Vol. 100/12, pp. 2366-2371, https://doi.org/10.2105/AJPH.2010.192245. [25]

Koushede, V. and R. Donovan (2022), "Applying Salutogenesis in Community-Wide Mental Health Promotion", in *The Handbook of Salutogenesis*, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-030-79515-3_44. [18]

New Zealand Initial Mental Health and Wellbeing Commission (2020), *Developing a Mental Health and Wellbeing Outcomes Framework Summary of what we heard through the co-define phase*, https://www.mhwc.govt.nz/assets/Outcomes-framework/Co-define-Report-on-Responses_Online-Version.pdf (accessed on 1 June 2022). [29]

Nielsen, L. et al. (2016), "High and low levels of positive mental health: are there socioeconomic differences among adolescents?", *Journal of Public Mental Health*, Vol. 15, pp. 37-49. [32]

OECD (2021), *A New Benchmark for Mental Health Systems: Tackling the Social and Economic Costs of Mental Ill-Health*, OECD Health Policy Studies, OECD Publishing, Paris, https://doi.org/10.1787/4ed890f6-en. [3]

OECD (2021), *COVID-19 and Well-being: Life in the Pandemic*, OECD Publishing, Paris, https://doi.org/10.1787/1e1ecb53-en. [5]

OECD (2021), *Fitter Minds, Fitter Jobs: From Awareness to Change in Integrated Mental Health, Skills and Work Policies*, Mental Health and Work, OECD Publishing, Paris, https://doi.org/10.1787/a0815d0f-en. [7]

OECD (2020), *How's Life? 2020: Measuring Well-being*, OECD Publishing, Paris, https://doi.org/10.1787/9870c393-en. [4]

OECD (2019), *Health at a Glance 2019: OECD Indicators*, OECD Publishing, Paris, https://doi.org/10.1787/4dd50c09-en. [40]

OECD (2015), *2015 OECD Recommendation on Integrated Mental Health, Skills and Work Policy*, https://www.oecd.org/els/emp/Flyer_MHW%20Council%20Recommendation.pdf (accessed on 13 March 2022). [6]

OECD (2013), *OECD Guidelines on Measuring Subjective Well-being*, OECD Publishing, Paris, https://doi.org/10.1787/9789264191655-en. [52]

OECD (2011), *How's Life?: Measuring Well-being*, OECD Publishing, Paris, https://doi.org/10.1787/9789264121164-en. [8]

OECD/European Union (2018), *Health at a Glance: Europe 2018: State of Health in the EU Cycle*, OECD Publishing, Paris/European Union, Brussels, https://doi.org/10.1787/health_glance_eur-2018-en. [41]

Patel, V. et al. (2018), "The Lancet Commission on global mental health and sustainable development", *The Lancet*, Vol. 392/10157, pp. 1553-1598, https://doi.org/10.1016/S0140-6736(18)31612-X. [15]

Payton, A. (2009), "Mental Health, Mental Illness, and Psychological Distress: Same Continuum or Distinct Phenomena?", *Journal of Health and Social Behavior*, Vol. 50/2, pp. 213-227, https://doi.org/10.1177/002214650905000207. [16]

Peterson, C. and M. Seligman (2004), *Character Strengths and Virtues: A Handbook and Classification*, American Psychological Association. [55]

Phillips, M. (2009), "Is distress a symptom of mental disorders, a marker of impairment, both or neither?", *World psychiatry : official journal of the World Psychiatric Association (WPA)*, Vol. 8/2, pp. 91-2. [45]

Queensland Government (2015), *The dual continuum of mental health and mental illness*, https://cspm.csyw.qld.gov.au/practice-kits/mental-health/overview-of-mental-health-1/the-dual-continuum-of-mental-health-and-mental-ill (accessed on 17 May 2022). [23]

Rehm, J. and K. Shield (2019), "Global Burden of Disease and the Impact of Mental and Addictive Disorders", *Current Psychiatry Reports*, Vol. 21/2, p. 10, https://doi.org/10.1007/s11920-019-0997-0. [2]

Reis, H. and S. Gable (2003), "Toward a positive psychology of relationships.", in *Flourishing: Positive psychology and the life well-lived.*, American Psychological Association, Washington, https://doi.org/10.1037/10594-006. [51]

Robinson, P. (2012), "Deep phenotyping for precision medicine", *Human Mutation*, Vol. 33/5, pp. 777-780, https://doi.org/10.1002/humu.22080. [26]

Roger T. Mulder (2008), "An Epidemic of Depression or the Medicalization of Distress?", *Perspectives in Biology and Medicine*, Vol. 51/2, pp. 238-250, https://doi.org/10.1353/pbm.0.0009. [46]

Routledge, K. et al. (2016), "Shared versus distinct genetic contributions of mental wellbeing with depression and anxiety symptoms in healthy twins", *Psychiatry Research*, Vol. 244, pp. 65-70, https://doi.org/10.1016/j.psychres.2016.07.016. [12]

Ryff, C. and C. Keyes (1995), "The structure of psychological well-being revisited.", *Journal of Personality and Social Psychology*, Vol. 69/4, pp. 719-727, https://doi.org/10.1037/0022-3514.69.4.719. [57]

Santini, Z. et al. (2022), "Higher levels of mental wellbeing predict lower risk of common mental disorders in the Danish general population", *Mental Health & Prevention*, Vol. 26, p. 200233, https://doi.org/10.1016/j.mhp.2022.200233. [27]

Santini, Z. et al. (2017), "The protective properties of Act-Belong-Commit indicators against incident depression, anxiety, and cognitive impairment among older Irish adults: Findings from a prospective community-based study", *Experimental Gerontology*, Vol. 91, pp. 79-87, https://doi.org/10.1016/j.exger.2017.02.074. [37]

Santini, Z. et al. (2020), "Predictors of high and low mental well-being and common mental disorders: Findings from a Danish population-based study", *European Journal of Public Health*, Vol. 30/3, pp. 503-509, https://doi.org/10.1093/eurpub/ckaa021. [33]

Snow, N. (2019), "Positive psychology, the classification of character strengths and virtues, and issues of measurement", *The Journal of Positive Psychology*, Vol. 14/1, pp. 20-31, https://doi.org/10.1080/17439760.2018.1528376. [54]

Solin, P. et al. (2019), "Predicting high and low levels of positive mental health among people living alone in the Lapland region of Finland", *Journal of Public Mental Health*, Vol. 18, pp. 214-223. [38]

Statistics Canada (n.d.), *Positive mental health and mental illness*, https://www150.statcan.gc.ca/n1/pub/82-003-x/2014009/article/14086-eng.htm (accessed on 17 May 2022). [19]

Steger, M. et al. (2006), "The meaning in life questionnaire: Assessing the presence of and search for meaning in life.", *Journal of Counseling Psychology*, Vol. 53/1, pp. 80-93, https://doi.org/10.1037/0022-0167.53.1.80. [50]

Stewart-Brown, S. et al. (2015), "Socioeconomic gradients and mental health: implications for public health", *British Journal of Psychiatry*, Vol. 206/6, pp. 461-465, https://doi.org/10.1192/bjp.bp.114.147280. [31]

Swiss Health Observatory (n.d.), *Psychischer Gesundheitszustand [Mental Health Status]*, https://www.obsan.admin.ch/de/gesundheitsthemen/psychische-gesundheit/psychischer-gesundheitszustand (accessed on 17 May 2022). [21]

Thoits, P. (2011), "Mechanisms Linking Social Ties and Support to Physical and Mental Health", *Journal of Health and Social Behavior*, Vol. 52/2, pp. 145-161, https://doi.org/10.1177/0022146510395592. [39]

Trent, D. (1992), "The promotion of mental health: fallacies of current thinking", in Trent, D. and C. Reeds (eds.), *Promotion of Mental Health*, Aldershot, Avebury. [13]

Van Lente, E. et al. (2012), "Measuring population mental health and social well-being", *International Journal of Public Health*, Vol. 57/2, pp. 421-430, https://doi.org/10.1007/s00038-011-0317-x. [36]

Varin, M. et al. (2020), "At-a-glance - An update on positive mental health among adults in Canada", *Health Promotion and Chronic Disease Prevention in Canada*, Vol. 40/3, pp. 86-91, https://doi.org/10.24095/hpcdp.40.3.04. [34]

Vinko, M. et al. (2022), "Positive mental health in Slovenia before and during the COVID-19 pandemic", *Frontiers in Public Health*, Vol. 10, https://doi.org/10.3389/fpubh.2022.963545. [35]

Wakefield, J. et al. (2007), "Extending the Bereavement Exclusion for Major Depression to Other Losses", *Archives of General Psychiatry*, Vol. 64/4, p. 433, https://doi.org/10.1001/archpsyc.64.4.433. [47]

World Health Organization (2022), *World mental health report: Transforming mental health for all*, https://www.who.int/publications/i/item/9789240049338 (accessed on 7 December 2022). [24]

World Health Organization (2021), *International Classification of Diseases - ICD -11 for Mortality and Morbidity Statistics*. [43]

World Health Organization (2019), *Mental health*, https://www.who.int/news-room/facts-in-pictures/detail/mental-health (accessed on 16 May 2022). [1]

World Health Organization (1948), *WHO Constitution Preamble*, https://www.who.int/about/governance/constitution (accessed on 5 April 2022). [9]

Notes

[1] As described later, mental ill-health in this report refers to diagnosable mental and behavioural conditions, as well as the transdiagnostic characteristic of general psychological distress. This terms mental health condition and mental disorder are used interchangeably in this report to refer to clinically significant symptoms of mental ill-health. Positive mental health covers psychological, emotional, and in some cases also social, relational and spiritual well-being. This report mainly focuses on the areas of positive mental health that have a strong overlap with the related concept of subjective well-being and that have been covered in-depth in the 2013 *OECD Guidelines of Subjective Well-being*, which define it as "good mental states, including all of the various evaluations, positive and negative, that people make of their lives, and the affective reactions of people to their experiences" (OECD, 2013[52]).

[2] For instance, the share of people at risk of depression in *How's Life? 2020* was reported only for European countries covered by the European Health Interview Survey (which is conducted only every five to six years), and information on negative affect balance (the share of the population reporting more negative than positive feelings and states) is currently sourced from the Gallup World Poll. Similarly, several of the surveys used to analyse inequalities in mental distress featured in the 2021 *A New Benchmark for Mental Health Systems* and *Fitter Minds, Fitter Jobs* workstreams were conducted before 2015 and use a variety of different (non-harmonised) instruments to measure distress.

[3] Current well-being is comprised of 11 dimensions: they relate to material conditions that shape people's economic options (income and wealth, housing, work and job quality) and quality-of-life factors that encompass how well people are (and how well they feel they are), what they know and can do, and how healthy and safe their places of living are (health, knowledge and skills, environmental quality, subjective well-being, safety). Quality of life also encompasses how connected and engaged people are, and how and with whom they spend their time (work-life balance, social connections, civic engagement). Resources for future well-being are expressed in terms of a country's investment in (or depletion of) different types of capital resources that last over time but that are also affected by the decisions taken (or not taken) today, and these include economic capital (man-made and financial assets), natural capital (stocks of natural resources, land cover, species biodiversity, as well as ecosystems and their services), human capital (skills and the future health of individuals) and social capital (social norms, shared values and institutional arrangements that foster cooperation) (OECD, 2020[4]).

[4] The indicator dashboard accompanying the OECD Well-being Framework differentiates between current well-being and the resources needed to sustain it, relying on different indicators for the two domains – hence, only premature mortality and obesity prevalence are included in the human capital indicator set (Exton and Fleischer, 2023[61]). However, people's physical and mental health, which are covered in other dimensions, influence their opportunities in later life and are conceptually within the scope of human capital.

[5] His broad view is also mirrored in the WHO's definition of health more broadly as "complete physical, mental and social well-being and not merely the absence of disease and infirmity".

[6] Various names for dual-continua models have been proposed, including the dual-factor model, two-factor, two-continua, the complete state model and complete mental health.

[7] Prevalence of mental health, or flourishing, was defined here as both symptoms of hedonia and positive functioning, and it is measured by six questions about positive affect, Ryff's scales of psychological well-being and Keyes' scales of social well-being (Keyes, 2005[11]).

[8] These studies were performed in clinical and non-clinical populations, over the entire life-course, and in Western and non-Western populations, and included studies specifically recruiting minority and at-risk groups.

[9] Both the Diagnostic and Statistical Manual of Mental Disorders (DSM) of the American Psychiatric Association and the WHO International Classification of Diseases (ICD) list a set of criteria that are needed for a diagnosis of a specific mental health condition to be met (World Health Organization, 2021[43]; American Psychiatric Association, 2013[42]). These criteria, which vary depending on the specific disorder, specify the nature and number of symptoms and the level of distress or impairment required, and are used to exclude cases where symptoms can be directly attributed to general medical conditions, such as a physical injury, or an expectable or culturally approved response to a common stressor or loss, such as the death of a loved one. Mainly used and developed in the United States by American psychiatry experts, the DSM is a specified classification system for mental disorders only, while the ICD is an overarching joint classification system for both physical and mental disorders. The first version of the DSM was published in 1952 and included 106 specific diagnoses. It has since been revised several times with the latest version (DSM-5) having been published in 2013, listing a total of 157 diagnoses and close to 300 disorders. A text revision (DSM-5-TR) was released in March 2022 that includes among other things updated diagnostic criteria and diagnostic codes, Prolonged Grief Disorder as new mental health condition, and considerations of the impact of racism and discrimination on mental disorders (American Psychiatric Association, 2022[64]). The ICD has a chapter (chapter F) devoted specifically to psychiatric disorders and is also regularly updated, with version 11 published in 2019. Although the two systems present minor differences, they are based on similar sets of rules and assumptions.

[10] The Institute for Health Metrics and Evaluation's burden of disease estimates for these mental health disorders are based on a wide variety of data sources and a set of complex assumptions regarding prevalence of a given disorder or risk factor and the relative harm it causes to quality of life and premature mortality.

[11] The way positive mental health is conceived, sometimes with greater focus and sometimes more broadly, is apparent in the way different government agencies across the OECD have operationalised the concept: the Canadian Positive Mental Health Surveillance Indicator Framework defines it as "a state of well-being that allows us to feel, think, and act in ways that enhance our ability to enjoy life and deal with the challenges we face" (Government of Canada, n.d.[62]), whereas the Finnish Institute for Health and Welfare describes it as "various levels of emotional (feelings), psychological (positive actions), social (relationships with others and society), physical (physical health and fitness) and spiritual (the sense that life has a meaning) wellbeing" (Finnish Institute for Health and Welfare, n.d.[63]).

2 Measuring population mental health: Tools and current country practice

A variety of tools are available for monitoring population mental health, ranging from administrative data to different types of survey questions. Although many OECD countries began collecting new or additional mental health data during COVID-19, official data producers were already active in this space well before the pandemic started. However, there is room for improvement by increasing the frequency of (survey) data collection, diversifying the types of indicators used to cover the full spectrum of mental health, and expanding the international harmonisation of existing measures. Here, data collectors could: (1) beyond screening tools focusing on symptoms of depression, expand use to those including symptoms of anxiety as outcome measures; (2) move towards collecting harmonised information on affective and eudaimonic aspects of positive mental health; and (3) explore using single-item questions on general mental health status across surveys.

The frequent collection of population-level data on mental health outcomes is important for identifying populations at-risk for mental ill-health, for determining which socio-economic and other factors shape (and are shaped by) people's mental health, and for designing effective prevention and promotion strategies. As outlined in Chapter 1, mental health is a multifaceted concept and exists beyond a binary distinction between the presence or absence of mental illness. Collecting data on both mental ill-health and positive mental health in population surveys and mental health assessments would yield a more complete picture of people's overall mental health and help to better understand the drivers and policy levers associated with improving it.

However, the current lack of (internationally) standardised data on population mental health makes it difficult to assess the efficacy of different policy approaches across disparate contexts; standardising outcome measures is the first step in facilitating such analysis. This chapter outlines the tools available to data collectors, gives an overview of current data collection practices across OECD countries and offers suggestions for which outcomes to prioritise in international harmonisation efforts.

An analysis of responses to a questionnaire sent to official data producers in OECD countries in 2022 shows that all member states that answered are already active in this space. Prior to the pandemic, almost all OECD members were already collecting information on mental health outcomes in both health interviews and general household surveys, as well as via administrative data. COVID-19 has sparked additional interest in measuring population mental health, with many public agencies and statistical offices adding items to both new and existing surveys.

These existing data collections demonstrate the interest in, and relevance of, population mental health outcomes in a national statistics context. Yet there is room for improvement in several areas: the frequency of data collection; greater data availability across the full spectrum of both negative and positive mental health outcomes; and better harmonisation of measures across countries to improve international comparability.

Indeed, prior to the pandemic most mental health data were collected by countries on surveys that ran every four to ten years. While many introduced high-frequency surveys with mental health modules in the first two years of COVID-19, it is currently unclear whether these surveys will continue to be implemented moving forward. Further, although all statistical offices collect data on mental ill-health – with a particular focus on common mental disorders – general psychological distress and depressive symptoms tend to be captured through standardised screening tools, whereas measures of experiencing anxiety are less harmonised across countries. Data collection efforts for other mental conditions – such as post-traumatic stress disorder (PTSD), bipolar disorder, eating disorders, etc. – and for other aspects of mental health – such as suicidal ideation and mental health-related stigma – remain very uneven across countries. When it comes to positive mental health, cross-country comparative data are mainly limited to measures of life evaluation. Other aspects, such as affect and eudaimonia, are much less frequently collected as outcome measures, and when they are, the tools used are less likely to be standardised across countries.

The results of the OECD questionnaire suggest that existing data collection efforts are not capturing the full range of mental health outcomes – missing aspects of both mental ill-health as well as positive mental health. In order to capture these outcomes and collect frequent information on mental health, data collectors in OECD member countries could: (1) beyond screening tools focusing on symptoms of depression, expand use to those including symptoms of anxiety as outcome measures; (2) move towards collecting harmonised information on affective and eudaimonic aspects of positive mental health; and (3) explore using single-item questions on general mental health status across surveys.

Which tools are available for measuring population mental health?

While Chapter 1 focused on relevant *types of outcomes* (covering both mental ill-health as well as positive mental health) for data collectors interested in mental health, this chapter focuses mainly on the *types of tools* that can be used to measure these.

The broad tool types discussed in this chapter – some of which are sourced from administrative data, but the bulk of which come from household surveys – range from long survey modules to a battery of question items to single questions. Some tools can be used to capture aspects of either mental ill-health or positive mental health, while others are used only for specific types of outcomes. Each type of tool has its own advantages and disadvantages, requiring data collectors to select among them, depending on the needs and constraints of their specific contexts. The different tools are described below in order to provide a common understanding of the categorisation used in this report.

The chapter annexes contain in-depth information for readers interested in further details. Annex 2.A provides an overview of which specific tools are collected by each country, along with sample question framing and answer options. Annex 2.B lists full details, including question wording and scoring recommendations, for the most commonly used standardised instruments. More detailed reflections on the statistical quality of mental health survey measures are addressed in Chapter 3.

Tools sourced from administrative data

Administrative data can contain information on the **use of mental health services**, **diagnoses of mental disorders in clinical settings**, as well as **cause of death data from suicide and substance abuse** (i.e. drug overdoses and alcohol abuse).

While all of these can be considered objective (i.e. not self-reported) and easy-to-collect proxies of mental ill-health, measurement challenges remain. For instance, measures of service use and medical diagnoses do not capture *population* outcomes, but rather only those who are willing and able to access health care services. Such measures can overestimate comparative levels or incidence rates in countries with good (and affordable) medical systems, awareness programmes and less stigma, where people are more likely to both seek and receive treatment. In addition, preventing ill-health necessitates tracking outcomes prior to, and following, engagement with the service sector. This report does not consider administrative statistics related to health care further, referring readers to (OECD, 2021[1]).

Data on causes of death due to suicide or substance abuse (which are commonly referred to as “deaths of despair” (Case and Deaton, 2017[2])) do capture mental ill-health outcomes at the population level. These measures can act as proxies for severe mental illness and addiction. While there are social and cultural reasons affecting suicidal behaviours – meaning that not all suicides are the direct result of a mental ill-health – living with mental health conditions does substantially increase the risk of dying by suicide (OECD, 2021[1]). However, the registration of suicide deaths is a complex procedure, affected by factors such as how intent is ascertained, who completes the death certificate, and prevailing norms and stigma around suicide, all potentially affecting the cross-country comparability of mortality records (OECD, 2021[1]).

A general limitation for all types of administrative data is that the additional socio-demographic data collected alongside are often limited to the age, sex, geographic region and potentially the race/ethnicity of the deceased. This constrains the ability to delve into the drivers of mental health and to identify relevant socio-economic, environmental and relational risk and resilience factors.

Tools sourced from household surveys

In contrast to administrative data, population surveys generally contain information on respondents’ material conditions (e.g. income, wealth, labour market outcomes, housing quality), quality of life (e.g.

physical health, educational attainment, environmental quality) and relationships (e.g. social connections, trust, safety). Population surveys can have a specific content focus, such as a health survey, or a more general scope, such as general social surveys. These surveys are conducted at the household level, with more in-depth modules on employment, health (including mental health), education, etc., administered to selected household members. Having a full range of well-being covariates is important to understand how mental health is impacted by, and how it in turn influences, other areas of people's life. Furthermore, tracking (and eventually achieving) equity in mental health outcomes requires disaggregation by important socio-demographic categories.

Tools that have been included in household surveys to assess specific mental health outcomes range from single-item questions to standardised batteries of items. A brief description of each can be found below, with full details in Annex 2.A and Annex 2.B.

- **Questions about previous diagnoses** – This refers to single-item questions about whether an individual has been diagnosed with a mental health disorder (e.g. major depressive disorder, generalised anxiety disorder, or other mental health conditions) by a health care worker, either in the past 12 months or over the course of his/her lifetime. These questions typically have yes/no answers and are not standardised across countries. For full details, see Table 2.6. Examples include:
  - "Have your mental health problems ever been diagnosed as a mental disorder by a professional (psychiatrist, doctor, clinical psychologist)? Yes / No".
  - "Have you EVER been told by a doctor or other health professional that you had ...Any type of depression? Read if necessary: Some common types of depression include major depression (or major depressive disorder), bipolar depression, dysthymia, post-partum depression, and seasonal affective disorder. Yes / No".
- **Questions about experienced symptoms** – This refers to single-item questions about symptoms of mental disorders experienced in the past 12 months or over the course of an individual's lifetime, without explicitly referring to a diagnosis by a medical professional. These questions typically have yes/no answers and are not standardised across countries. For full details, see Table 2.7. Examples include:
  - "During the past 12 months, have you had any of the following diseases or conditions? Depression ("Yes / No").
  - "Have you ever suffered from chronic anxiety? ("Yes / No").
  - "Do you have a mood disorder? Yes / No".
- **Questions about suicidal ideation and suicide attempts** – These are (usually) single-item questions about a respondent's experience of suicidal ideation, self-harm behaviours or suicide attempts. These questions typically have yes/no answers and are not standardised across countries. Recall periods refer to an individual's lifetime, the last 12 months, the past two weeks, or "during COVID". For full details, see Table 2.8. Examples include:
  - "Have you seriously contemplated suicide since the COVID-19 pandemic began? Yes/No".
  - "Sometimes people harm themselves on purpose but they do not mean to take their life. In the past 12 months, did you ever harm yourself on purpose but not mean to take your life? Yes/No".
  - "Have you ever attempted suicide? Yes/No".
  - "Did you stay in a hospital overnight or longer because you tried to kill yourself? Yes/No".
- **Questions about general mental health status** – These refer to single-item questions on how respondents rate their mental health overall, and thus capture both components of ill-health and positive mental health. Questions are not standardised across countries and differ in terms of question wording, response options and recall period. For full details, see Table 2.9. Examples include:

- "In general, how is your mental health? Excellent / Very good / Good / Fair / Poor".
- "Has your mental health/well-being been affected by the Covid-19 pandemic during the last 12 months?"
- "On a scale from 1 to 10 can you indicate to what extent you are satisfied with your mental health? A score of 1 refers to completely dissatisfied and a 10 to completely satisfied".
- "Does your mental state interfere with your daily life at work? your family life? Yes / No".

- **Positive mental health indicators** – This refers to questions pertaining to the various aspects of positive mental health: life evaluation, affect (summary affect scales, and batteries of questions on positive, negative or mixed affect), eudaimonia (questions about quality of life, whether life is worthwhile or meaningful), as well as standardised positive mental health composite scales (combining different dimensions of positive mental health, prioritising positive over negative affect, and sometimes adding a social well-being component). In some instances, positive mental health indicators are single-item questions that vary across countries and surveys, while in others they are standardised batteries of questions. Standardisation across countries varies, with only life evaluation questions and positive mental health composite scales being consistently phrased. For full details, see Table 2.10. Specific question item phrasing and scoring suggestions for standardised composite scales can be found in Annex 2.B.
- **Screening tools** – These refer to multi-item instruments designed to screen respondents for symptoms (rather than for diagnoses) of mental health conditions. These tools were initially developed in clinical settings to screen for common mental disorders to identify individuals who may be at risk and to flag them for further screening and potential diagnosis. They can be interviewer-led or self-administered and focus either on general psychological distress or on specific mental health conditions such as major depressive disorder, generalised anxiety disorder (and sometimes a combination of the two), alcohol use disorder, post-traumatic stress disorder, eating disorders and so on. These tools are considered "validated" in that they have been psychometrically tested for their validity (against the gold standard of structured interviews or diagnoses), sensitivity (the probability of correctly identifying a patient with the condition) and reliability (the measures produce consistent results when an individual is interviewed under a given set of circumstances) (refer to Chapter 3 for an extended discussion of statistical quality). A wide variety of screening tools are available, ranging from very short screeners of two items to longer instruments covering 20 items or more. The focus of questions varies between screening tools: all cover the frequency of experiencing (mostly negative) affect (i.e. feeling low, feeling nervous, feeling worthless), with some also including somatic symptoms (i.e. changed appetite, trouble sleeping) and/or functional impairment due to emotional distress (e.g. disturbance in daily activities, not being able to concentrate, not being able to stop worrying). Screening tools also differ in terms of reference period for symptoms, ranging from the past week to the past month; however, none are able to measure lifetime prevalence. Given these differences between screening tools, they are therefore not always directly comparable and should not be used interchangeably for international comparisons. Item scores are typically summarised in a summary index, with the final score being used either as a continuous measure of mental ill-health or to assess the risk of a common mental health conditions using a validated cut-off score. For full details, refer to Table 2.5. Exact question item wording and scoring recommendations for the most frequently used screening tools can be found in Annex 2.B.
- **Structured interviews –** Structured interviews are considered the gold standard for measuring mental disorders (often both on a lifetime and 12-month basis). They provide a standardised assessment based on the internationally agreed definitions and criteria of recognised psychiatric classification systems and have strong diagnostic reliability and psychometric properties to determine whether or not a respondent has the condition of interest (Mueller and Segal, 2015[3]; Burger and Neeleman, 2007[4]).[1] They are administered by trained interviewers, with close-ended

and fully scripted questions and standardised scoring of responses (Ruedgers, 2001[5]). Structured interviews approximate assessments conducted by mental health professionals and in this way can identify populations at risk for mental health conditions even if these individuals have not been diagnosed by a health care professional. For additional information on the most commonly used structured interview, the Composite International Diagnostic Interview (CIDI), see Table 2.4 and Annex 2.B.

- **Additional mental-health related topics** – This category refers to questions on any other relevant topics, including the use of mental health medication and services, the mental health of children and young people in the household, loneliness and stress, resilience and self-efficacy, attitudes towards mental health including stigma and literacy, and questions on unmet needs. For additional information, see Table 2.11.

Trade-offs between tool types

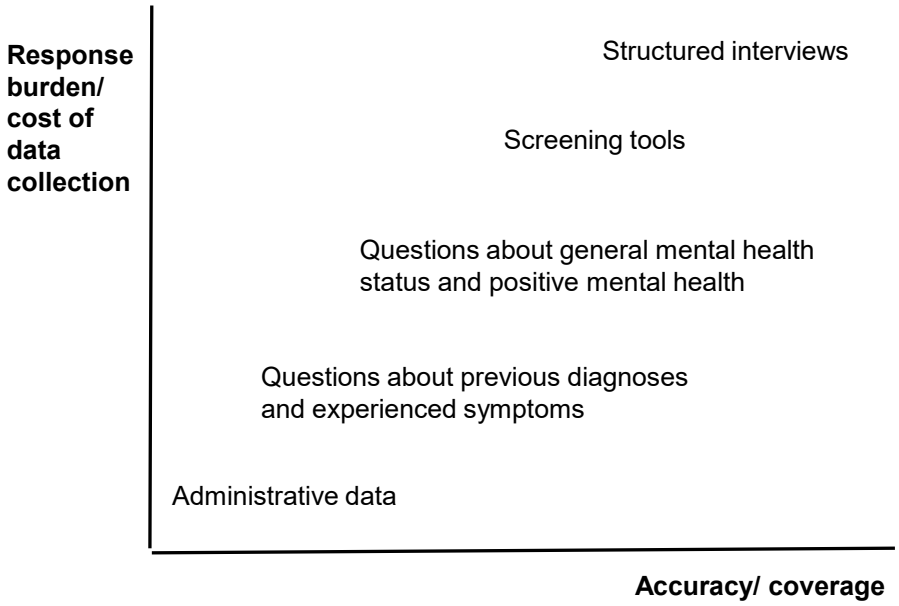
All tools imply trade-offs in terms of response burden/ease and cost of data collection, accuracy and coverage (Figure 2.1, Table 2.1). Response burden is a direct function of how much time an individual needs to spend to provide information on their mental health status and how much stress is caused by providing this information. Accuracy refers to the sensitivity of a tool in correctly identifying a person with a mental health condition, whereas coverage entails whether the measure in question is applied to the full (adult) population.

By way of illustration, administrative data have a low response burden: they do not require answers from individual respondents and are routinely collected within a country's data infrastructure. Yet statistics on deaths of despair focus only on the extreme end of mental ill-health and are further complicated by the fact that not all deaths of despair may be the culmination of a mental disorder. Furthermore, unlike household surveys, only those who were in contact with the health care system are captured by administrative records of diagnoses in a clinical setting.[2]

For household surveys, both the response burden and accuracy increase the longer and more specific a tool is: whereas single questions about experienced symptoms or a person's general mental health status are short and easy to answer, they do not consider the nature or severity of symptoms, or the type of mental health condition, and have not been benchmarked against diagnostic criteria. Screening tools have been validated against the gold standard of structured interviews and are, depending on the specific tool used and the number of items covered, still relatively low cost in terms of response burden. However, they do not constitute a diagnosis from a health care professional and can only identify people likely *at risk* of disorders. Screening tools are validated against clinical diagnoses, and are thus designed to maximise likeness to diagnostic interviews to the extent possible. Still, when calibrating tools and cut-off scores, there is a trade-off between sensitivity (correctly identifying the presence of a mental health condition) and specificity (correctly noting the *absence* of a mental health condition), and researchers often prioritise the former rather than the latter, leading to slight overestimates by design (see Box 3.3 and Section 3.3.1 for a more detailed discussion). Finally, the majority of tools included in both household surveys and administrative data focus on mental ill-health; the only exceptions are household survey questions about general mental health and positive mental health.

The difference in question framing and item length – between structured interviews, screening tools and single-item questions on experienced symptoms or received diagnoses – can lead to different estimates of prevalence for the same reported outcome measure (Box 2.1). This speaks to the need for the standardisation of tool type (and transparency about which tool was used) when comparing outcomes across countries, over time and across population groups: i.e. mixing types of tools when commenting on outcomes like "share at risk for depression" or "share at risk for psychological distress" can lead to different estimates because of measurement differences, rather than because of differences in underlying mental health status (refer to Chapter 3 for an extended discussion of these themes).

Figure 2.1. Trade-off between response burden and accuracy for mental health measurement tools



Source: Adapted from a presentation given by Statistics Canada at the OECD conference "Well-being and mental health – towards an integrated policy approach" in December 2021.

Table 2.1. Advantages and limitations of different tools to measure mental health

Tool	Advantages	Limitations
Administrative data (deaths of despair from suicide, drug overdose, alcohol abuse; diagnoses of common mental disorders in clinical care settings)	• No response burden for individuals • Possibility to link across other administrative data (e.g. health system quality) • Less costly and more readily available than other types of data • Clinical care data can provide some insight into lifetime and specific time period (e.g. past 12 months) prevalence estimates for a range of ill-health conditions when other data sources are not available	• Captures only those who sought treatment, were correctly coded by a health professional and are part of the reporting database • "Cause of death" data need to be correctly coded, do not account for suicide attempts or substance abuse not leading to death, and only capture the extreme end of mental ill-health • Often difficult, or even impossible, to interpret (without supplemental information) whether changes in diagnostic rates are driven by changes in underlying prevalence of mental health conditions or by other factors such as changes to affordability or accessibility of care, changes in help-seeking behaviour, etc. • Limited contextual information on well-being covariates
Household surveys: **questions about previous diagnoses**	• Relatively easy to understand for respondents • Minimal response burden (usually a single binary question) • Can provide both lifetime and specific time period (e.g. past 12 months) prevalence estimates for a range of ill-health conditions	• Captures only those who sought treatment and were diagnosed by a health professional • Evidence that these questions lead to social desirability bias and higher rates of refusal and non-response (see Chapter 3) • Limited contextual information on the nature and severity of symptoms
Household surveys: **questions about experienced symptoms of mental health conditions**	• Minimal response burden (usually a single binary question) • Can provide both lifetime and specific time period (e.g. past 12 months) prevalence estimates for a range of ill-health conditions	• Potential for confusion for respondents in terms of whether the question refers to an actual diagnosis or their self-assessment, though evidence suggests this type of tool is closely related to questions about previous diagnoses by health professionals • Limited contextual information on the nature and severity of symptoms
Household surveys: **questions on general mental health status**	• Relatively easy to understand for respondents • Minimal response burden (usually a single	• Over-reporting of true prevalence – not a complete assessment or an actual diagnosis, does not consider symptoms

Tool	Advantages	Limitations
	question) • Captures a respondent's global evaluation of their mental state, and hence both ill-health and positive aspects	• Has not been validated against structured interviews or other diagnostic tools, no established threshold in the tools as to what constitutes at-risk respondents • Generally less of an existing evidence base, though available studies suggest this to be a useful measure • Limited contextual information on the nature and severity of symptoms or the type of mental health condition
Household surveys: **indicators of positive mental health**	• Relatively easy to understand for respondents • Minimal response burden (usually single or limited-item questions) • Focus on psychological and emotional well-being or flourishing • International measurement guidance exists (e.g. *OECD Guidelines on Measuring Subjective Well-being*)	• No reference point of what (true and/or desired) prevalence should be • Recall period for questions typically ranges from day prior to past 4 weeks; cannot provide lifetime estimates
Household surveys: **screening tools**	• Easy to administer and reduced response burden compared to structured interviews • Have been validated against structured interviews or other diagnostic tools • Can capture undiagnosed conditions	• Over-reporting of true prevalence – not a complete assessment or an actual diagnosis • Recall period for questions typically ranges from day prior to past 4 weeks; cannot provide lifetime estimates
Household surveys: **structured interviews**	• Approximates true prevalence – near gold standard • Can capture undiagnosed conditions • Extensive contextual information of the respondents' lives can be taken into account	• Very complex to develop and administer, including interviewer training • Many questions for people who have symptoms • Lack of survey measurement tools available to map to most up-to-date diagnostic guidelines (DSM-5)

Source: Adapted from a presentation given by Statistics Canada at the OECD conference "Well-being and mental health – towards an integrated policy approach" in December 2021.

Box 2.1. Prevalence rates vary depending on the measurement tool used

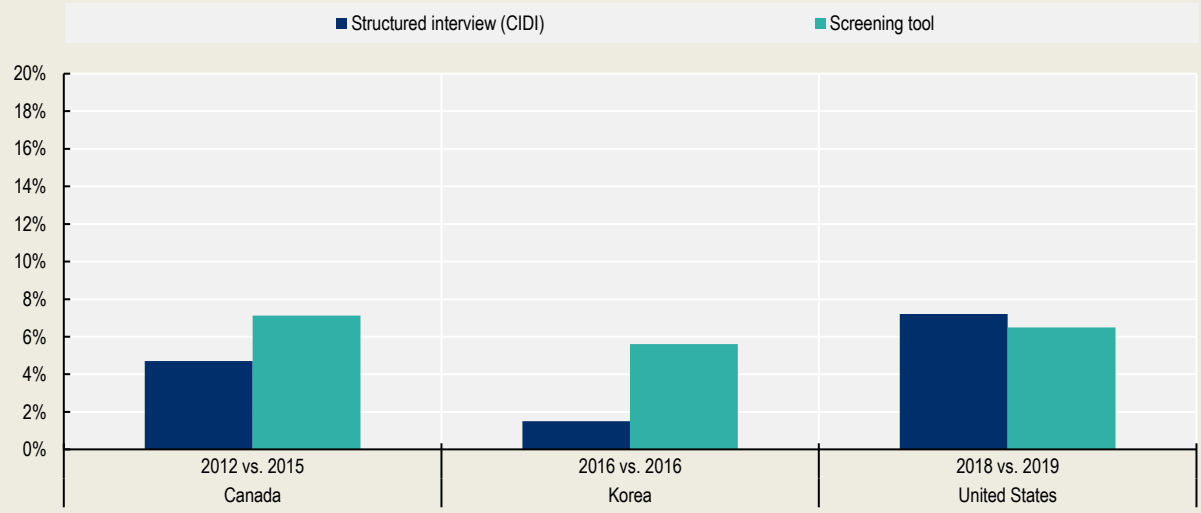
Prevalence rates for specific mental health conditions will vary – at times substantially – depending on the type of tool used to create the estimate. Screening tools are likely to *overstate* population level prevalence of mental disorders by design. They were developed in clinical settings to identify individuals at risk for common mental disorders, who can then be flagged for further observation and actual diagnoses – some of whom may not end up being diagnosed or needing further treatment (National Academies of Sciences Engineering and Medicine, 2021[6]; Topp et al., 2015[7]). In contrast, questions that require individuals to report whether they have been diagnosed with a mental disorder by a health care professional in the past, or currently live with a specific disorder, focus on those in touch with the health care system and are therefore likely to *understate* population prevalence.

On the first point, Figure 2.2 below shows that screening tools may overestimate population prevalence as compared to structured interviews. The figure shows national estimates of the same outcome measure – prevalence for major depressive disorder (MDE) – in three OECD countries as measured by CIDI, a structured interview, and by the Patient Health Questionnaire (PHQ), a screening tool. (The version of the PHQ varies by country: PHQ-9 in Canada and Korea, PHQ-2 in the United States. Refer to Annex 2.B for the specific items included in each iteration.) While both the CIDI and screening tools are used in different surveys within each country, implying that care should be taken in making direct comparisons, generally prevalence of MDE as measured by the CIDI is lower than that measured through screening tools. The exception is the United States, which also shows the smallest difference between the estimates. This may

be in part because many mental health survey tools were first developed, and subsequently extensively validated, in the United States, making the calibrations between different tools more precise.

Figure 2.2. Screening tools typically show greater prevalence of major depressive disorder than do structured interviews

Prevalence of major depressive episodes (MDE), over the past 12 months vs. past 2 weeks, as estimated by CIDI and screening tools (PHQ)



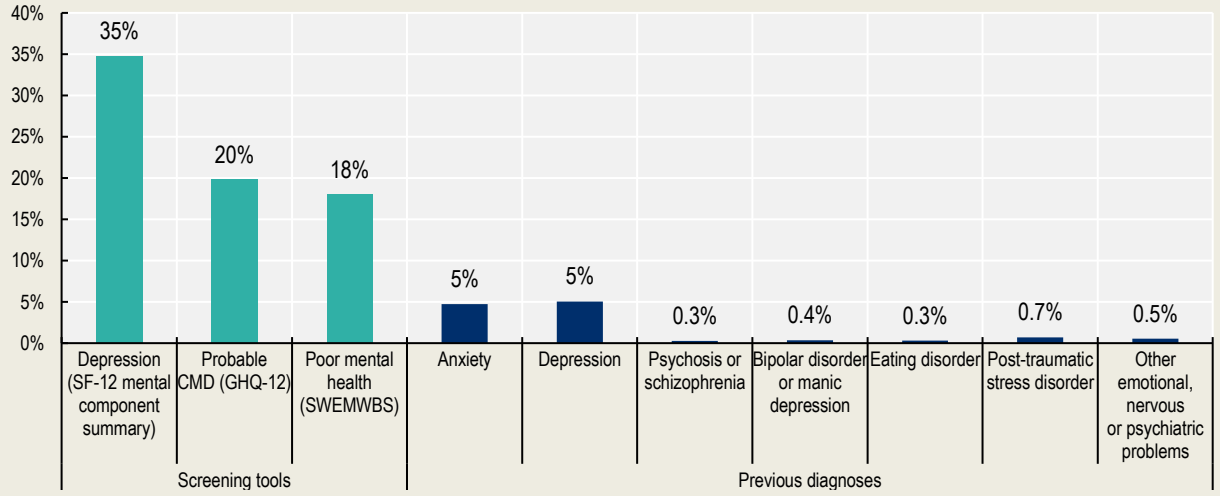
Note: For all three countries, the structured interview used is the CIDI, which is used to measure the prevalence of Major Depressive Episodes (MDE) over the past 12 months. In Korea, these estimates are adjusted for age and sex. In Canada and the United States, these estimates are nationally representative for the 15+ and 18+ population, respectively. The validated screening tool used by Canada is the PHQ-9 (MDE defined as having a score >= 10); the PHQ-9 is used by Korea (being at risk for depression is defined as having a score >= 10; although not described by KOSIS, Korea's statistical service, as a risk for MDE, this same scoring convention is used by Canada to measure MDE); and the PHQ-2 is used by the United States (symptoms of a depressive disorder are defined as having a score >= 3). The PHQ-9 and PHQ-2 both have a reference period of the past 2 weeks. For the United States, the PHQ-2 measures the share with symptoms of a depressive disorder, rather than experience of MDE. Refer to Annex 2.B for more information on individual screening tools.
Source: Structured interview data for Canada come from Statistics Canada (2013[8]), *Canadian Community Health Survey: Mental Health, 2012*, The Daily, https://www150.statcan.gc.ca/n1/daily-quotidien/130918/dq130918a-eng.htm; PHQ-9 data for Canada are derived from Dobson, K. et al. (2020[9]), "Trends in the prevalence of depression and anxiety disorders among Canadian working-age adults between 2000 and 2016", *Health Reports*, Vol. 31/12, pp. 12-23, https://doi.org/10.25318/82-003-X202001200002-ENG; Structured interview data for Korea come from KOSIS (n.d.[10]), *Annual prevalence of mental disorders (adjusted for sex and age)* (database), Korean Statistical Information Service, https://kosis.kr/statHtml/statHtml.do?orgId=117&tblId=TX_117_2009_HB027&conn_path=I2; PHQ-9 data for Korea come from KOSIS (KOSIS, n.d.[11]), *Depressive disorder prevalence* (database), National Health and Nutrition Survey, Korean Statistical Information Services, https://knhanes.kdca.go.kr/knhanes/sub01/sub01_05.do#none; Structured interview data for the United States come from SAMHSA (2019[12]), *Key Substance Use and Mental Health Indicators in the United States: Results from the 2018 National Survey on Drug Use and Health* (database), Substance Abuse and Mental Health Services Administration, Rockville, MD, https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHNationalFindingsReport2018/NSDUHNationalFindingsReport2018.pdf; PHQ-2 data for the United States come from the National Center for Health Statistics (2021[13]), *Estimates of Mental Health Symptomatology, by Month of Interview: United States, 2019* (database), U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, https://www.cdc.gov/nchs/data/nhis/mental-health-monthly-508.pdf.

StatLink https://stat.link/q5bvk9

On the second point, Figure 2.3 shows that, based on answers to screening tools, the share of the population reporting ever having received a diagnosis for a given mental disorder is much lower than the share deemed to be at risk for poor mental health conditions; this is often a function of affordability and access to health care, along with stigma and mental health illiteracy affecting health-seeking behaviours.

The left-hand side of the figure displays the share of respondents who are at risk for psychological distress or low levels of positive mental health, including: (1) those at risk for depression, as defined by a scoring convention of the Short Form-12 mental health summary component (SF-12); those at risk for a probable common mental disorder, as measured by the General Health Questionnaire-12 (GHQ-12); and (3) those who have poor mental well-being, as defined by a scoring convention of the Short Warwick–Edinburgh Mental Wellbeing Scale (SWEMWBS). (Refer to Table 2.5 and Table 2.10, along with Annex 2.B, for more information on the three tools.) The right-hand side of the figure shows the share of respondents who report having ever received a diagnosis for a range of specific mental health conditions.

Figure 2.3. The share of those reporting a diagnosis of a mental health condition is much lower than the share identified as experiencing psychological distress by screening tools



Note: Scoring information for each of the screening tools included: risk for depression is defined as having a score <= 45 on the transformed SF-12 mental health component composite scale, where 0 indicates worst mental health and 100 best possible mental health; risk for a probable common mental disorder (CMD) is defined as having a score >= 4 on the GHQ-12, as used in (Woodhead et al., 2012[14]); poor mental health is defined as having a SWEMWBS score more than one standard deviation below the sample average. Refer to Annex 2.B for more information on individual screening tools.

Source: OECD calculations based on University of Essex, Institute for Social and Economic Research (2022[15]), *Understanding Society: Waves 1-11, 2009-2020 and Harmonised BHPS: Waves 1-18, 1991-2009* (database), 15th Edition, UK Data Service, SN: 6614, http://doi.org/10.5255/UKDA-SN-6614-16, from wave 10 only (Jan 2018 – May 2020).

StatLink https://stat.link/9lqxu4

Which population mental health data are OECD countries already collecting?

In February and March of 2022, 37 of 38 OECD countries provided answers to a questionnaire designed by the OECD Secretariat to better understand what OECD countries are doing in terms of measuring mental health outcomes.[3] The questionnaire covers the statistical tools used (questions about diagnoses, experienced symptoms, screening tools and structured interviews) and outcomes covered (mental ill-health, positive mental health and other related topics, including loneliness, stress, attitudes towards mental health, etc.). A discussion of mental health data related to service use and access to care is set out in *A New Benchmark for Mental Health Systems* (OECD, 2021[1]), and this new round of surveying seeks to build upon existing work by primarily focusing on mental health *outcomes*, rather than on service use or access to care, and in particular on outcomes that could be measured through household surveys rather than administrative data.

All OECD countries already collect both administrative and survey data on population mental health

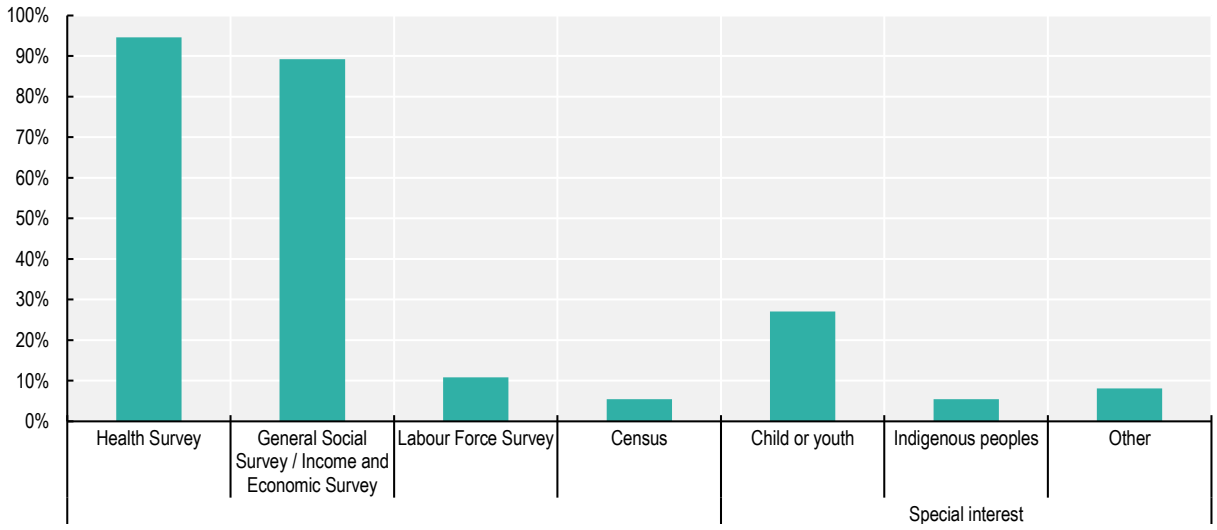
All OECD countries collect mortality statistics on causes of death, including from suicides rates as well as deaths from alcohol and drug overdoses. Statistics on causes of deaths are typically collected by hospitals or health care providers, while police authorities report deaths from suicides. The OECD already regularly publishes statistics for its member countries on both deaths from suicide and other types of deaths of despair (OECD, 2020[16]; OECD, 2021[17]).[4]

Administrative data on mental health go beyond death records. Hospital discharge registries that, depending on the country, may cover the length of hospitalisation and discharges by field of medical specialisation were mentioned by a number of countries, including Canada, Chile, Hungary, Italy, Slovenia, Switzerland and Türkiye. Some countries, including Spain and the United Kingdom, collect care or clinical care data to measure prevalence and incidence of specific behavioural disorders. The Swedish Social Insurance Agency also collects data on causes of work absences, with a special category for sick leave following a psychiatric diagnosis. Finally, a handful of countries collect administrative data on psychiatric medication. For example, in France the *Agence nationale de sécurité du médicament* (ANSM) publishes data on psychotropic drugs delivered to outpatients; Statistics Netherlands provides data on dispensed medicines, including those related to mental health conditions as determined by ATC (anatomical therapeutic chemical) coding; Australia collects administrative data on dispensed medications covered under the Pharmaceutical Benefits Scheme; and the Slovenian National Institute of Public Health (NIJZ) hosts data on prescription drug claims, including for mental health-related drugs.

In addition, all OECD countries that responded to the questionnaire reported collecting population-wide data on mental health outcomes through household surveys, already prior to the COVID-19 pandemic. While much of these data are collected through health interviews, 89% of countries reported also collecting mental health data in general social surveys (Figure 2.4). Some data on mental health are also collected through labour force surveys and special modules of the national census. Some countries also reported collecting mental health data in special surveys that focus on sub-populations, including Indigenous peoples, those in the criminal justice system and young people (see Box 2.2 for more information on the latter).

Figure 2.4. The majority of OECD countries report measuring population mental health in both health and general social surveys

Share of OECD countries that responded to a survey about population mental health



Note: Results are shown for all OECD countries except Estonia, which did not participate in the questionnaire.
Source: Responses to an OECD questionnaire sent to national statistical offices in January 2022.

StatLink https://stat.link/40bmf6

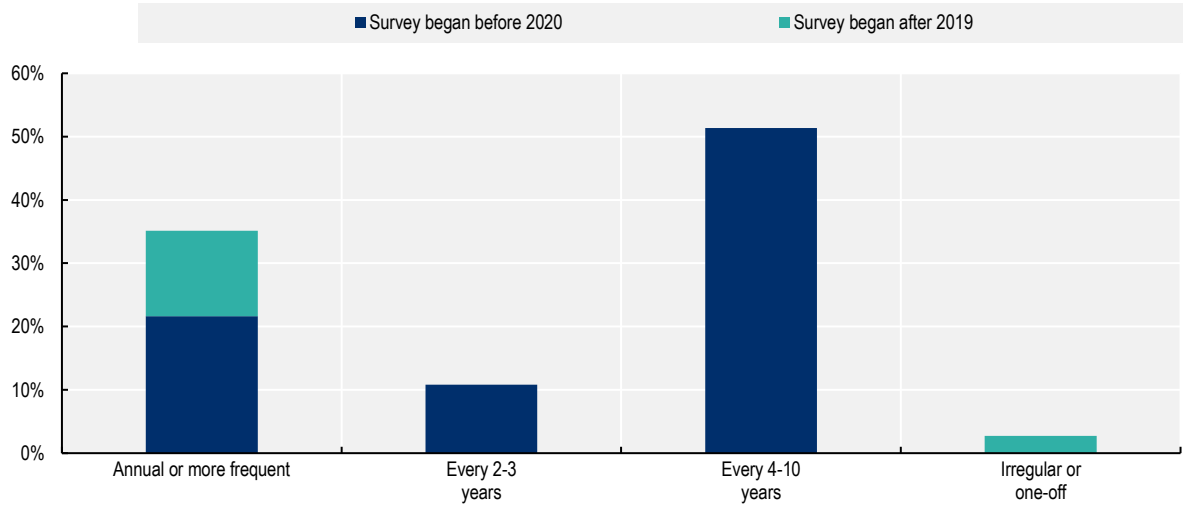
Many countries have launched surveys with mental health content since the onset of COVID-19, but it is unclear whether these will continue in the future

The pandemic has put mental health high on the national agenda for many OECD countries. As a result, most countries that answered the OECD questionnaire reported having ramped up data collection efforts on mental health in the months and years since March 2020. Around 68% of OECD countries reported collecting additional mental health data during the pandemic, either through new stand-alone surveys (43%) or by adding mental health and COVID-19 modules to existing surveys (35%) (see Table 2.3).[5] Many of these new surveys are high-frequency, interviewing respondents weekly, biweekly, monthly or quarterly. However, it is unclear whether these surveys will continue in the future, or continue with the same frequency. Indeed, some COVID-specific surveys have already been discontinued by countries, while others that started off as weekly or monthly have since become less frequent (biweekly or quarterly).

Before 2020, only 22% of countries collected mental health data on surveys that ran annually or more frequently, and 11% on surveys that ran every two to three years. Returning to business as usual prior to the pandemic would mean that over half (51%) of countries collect mental health data every four to ten years. Such large gaps between survey rounds make it more difficult to track changes at the population-level (which as has been seen during the COVID-19 pandemic were sensitive to periods of intensifying COVID-19 deaths and strict confinement measures) and craft policy interventions accordingly.

Figure 2.5. Many OECD countries collect mental health data infrequently, with over half reporting four-to-ten-year lags between survey rounds

Share of OECD countries that responded to a survey about population mental health



Note: This figure considers only the most frequently run survey per country, rather than the full set of surveys containing mental health data that countries report. It thus shows the highest degree of frequency for which mental health are available, per country. Results are shown for all OECD countries except Estonia, which did not participate in the questionnaire.
Source: Responses to an OECD questionnaire sent to national statistical offices in January 2022.

StatLink https://stat.link/mboi94

Box 2.2. Initiatives to collect data on mental health for children and youth

The mental health of young people suffered dramatically during the COVID-19 pandemic (OECD, 2021[18]; OECD, 2021[19]), and a number of OECD countries launched campaigns focusing on youth mental health in 2021 and 2022 to help combat increasing rates of suicide, reported anxiety, depression and general psychological stress (HHS, 2021[20]; Chile, 2021[21]; Santé Publique France, 2021[22]). The results from the OECD questionnaire show that, although the pandemic may have underscored the importance of focusing on young people, many OECD countries were already implementing child or youth-specific surveys with mental health modules (Table 2.2).

The measurement of child and youth mental health differs from that of adults in several ways. Some surveys use the same tools for children and adults – questions about previous diagnoses, standardised composite scales such as the WHO-5, negative affect questions – however, there are also some youth-specific validated screening tools. A number of countries answering the OECD questionnaire reported using the Strengths and Difficulties Questionnaire (SDQ), a behavioural screening tool for children and youth aged three to 16, or the Development and Well-Being Assessment (DAWMA), to screen for psychiatric diagnoses for children starting at age of two. Child and youth surveys often include modules covering behavioural and emotional issues, adverse childhood experiences, positive childhood experiences and substance use/abuse, and can contain questions that are posed to children, parents or teachers (Table 2.11). Some surveys also cover previous diagnoses of attention deficit hyperactivity disorder (ADHD) or autism spectrum disorder (ASD).

Table 2.2. Many countries have introduced child and youth surveys, or survey modules, with a mental health focus

Country	Survey
Australia	Australian Child and Adolescent Survey of Mental Health and Wellbeing
Canada	Canadian Health Survey of Children and Youth (CHSCY)
Germany	Study on the Health of Children and Adolescents in Germany (KiGGS)
Italy	Quality of Life in Children and Adolescents*
Luxembourg	Youth Survey Luxembourg
United Kingdom	Mental Health of Children and Young People Surveys
United States	Youth Risk Behavior Survey (YRBS) National Health Interview Survey (NHIS)†
Denmark, Ireland, Italy, Latvia, Luxembourg, Slovenia, Sweden	Health Behaviour in School-Aged Children (HBSC)

Note: The HBSC is a school-based survey, not a household survey. * indicates the survey was introduced following the start of the pandemic (post-March 2020). † The NHIS includes the Strengths and Difficulties Questionnaire (SDQ) in the child component of the rotating core module. Results are shown for all OECD countries except Estonia, which did not participate in the questionnaire.
Source: Responses to an OECD questionnaire sent to national statistical offices in January 2022.

Table 2.3. Over half of OECD countries reported increasing the collection of mental health data during the COVID-19 pandemic

Country	*Stand-alone COVID survey*	*COVID module added to existing survey*	*Any COVID-related survey*
Australia	●		●
Austria			
Belgium	●		●
Canada	●		●
Chile	●		●
Colombia	●		●
Costa Rica		●	●
Czech Republic			
Denmark			
Finland		●	●
France	●	●	●
Germany	●	●	●
Greece			
Hungary			
Iceland		●	●
Ireland	●		●
Israel	●		●
Italy		●	●
Japan			
Korea	●		●
Latvia			
Lithuania			
Luxembourg	●		●
Mexico	●		●
Netherlands		●	●
New Zealand		●	●

Country	*Stand-alone COVID survey*	*COVID module added to existing survey*	*Any COVID-related survey*
Norway		●	●
Poland			
Portugal			
Slovak Republic			
Slovenia	●		●
Spain		●	●
Sweden	●	●	●
Switzerland	●		●
Türkiye			
United Kingdom		●	●
United States	●	●	●

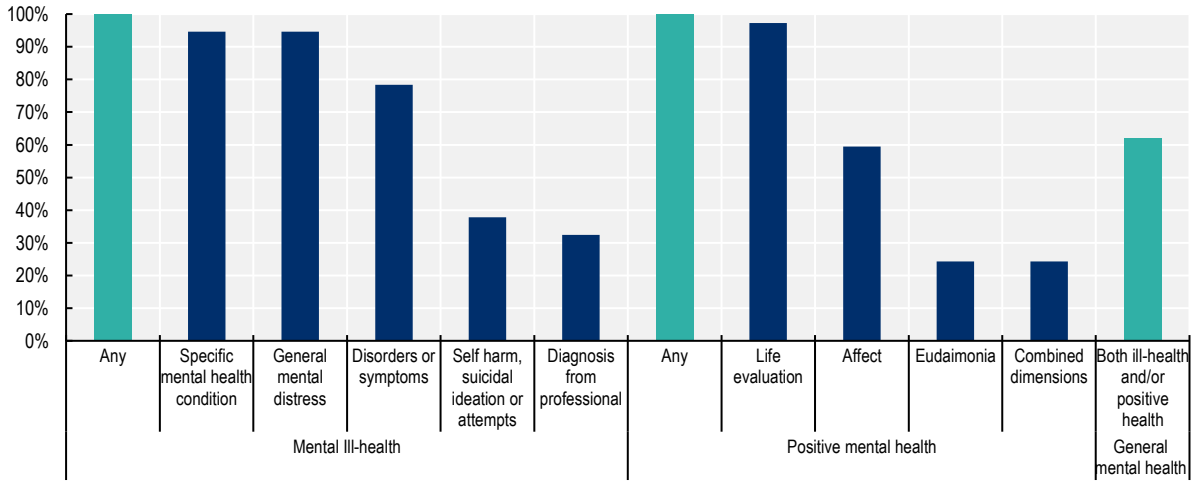
Note: Results are shown for all OECD countries except Estonia, which did not participate in the questionnaire.
Source: Responses to an OECD questionnaire sent to national statistical offices in January 2022.

The focus of household surveys is mainly on mental ill-health

All OECD countries collect data on both mental ill-health and positive mental health outcomes. For the former, there is much variety in terms of both the tools used and outcomes measured, whereas for the latter cross-country comparative data are mainly limited to measures of life evaluation (Figure 2.6); 59% of countries reported collecting data on affect, and only 24% on eudaimonia.

Figure 2.6. All OECD countries reported collecting data on mental ill-health and positive mental health, with the latter mostly focused on life evaluation

Share of OECD countries that responded to a survey about population mental health which report collecting data on various population mental health outcomes



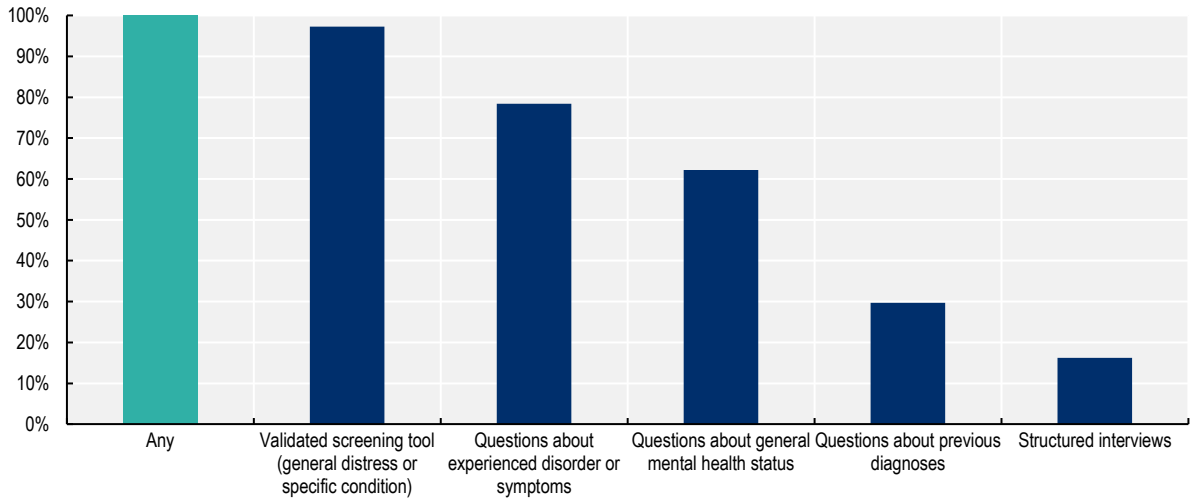
Note: Results are shown for all OECD countries except Estonia, which did not participate in the questionnaire. Note that the question collected during the EU-SILC 2013 ad hoc well-being module, on the extent to which respondents feel that their life is worthwhile, was not included in this figure given that the question was removed from subsequent well-being modules.
Source: Responses to an OECD questionnaire sent to national statistical offices in January 2022.

StatLink https://stat.link/cghuny

Mental ill-health outcome measures are captured through a variety of tools. The two tools most often reported by countries are screening tools and questions about experienced symptoms or disorders (either general or specific), with 97% and 78% of countries reporting using these types of tools in household surveys, respectively (Figure 2.7). Over half of countries (62%) ask single questions about people's general mental health status. Many fewer countries report collecting data on previous diagnoses in household surveys (30%) or in structured interviews (16%).

Figure 2.7. Screening tools and questions about experience of symptoms and disorders are the most common mental ill-health tools reported by countries

Share of OECD countries that responded to a survey about population mental health that measure mental ill-health by each type of tool



Note: Results are shown for all OECD countries except Estonia, which did not participate in the questionnaire.
Source: Responses to an OECD questionnaire sent to national statistical offices in January 2022.

StatLink https://stat.link/5twjx3

General psychological distress and symptoms of depression tend to be captured by standardised screening tools, whereas measures of experiencing anxiety are often not harmonised across countries

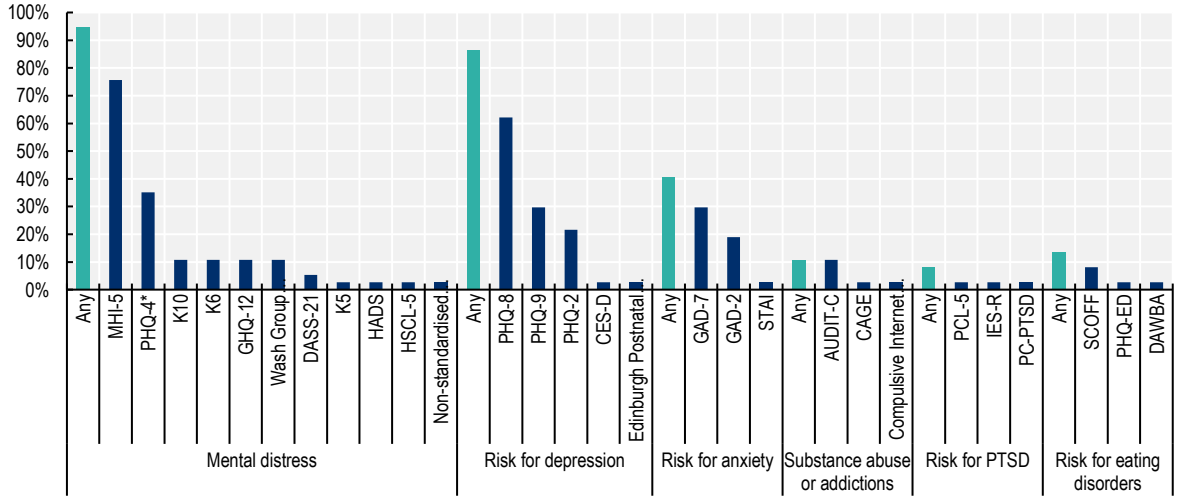
Within the continuum of mental ill-health, existing measurement initiatives focus more on some forms of mental health issues than on others. Anxiety and depressive disorder are the most common mental health conditions affecting people in OECD countries (OECD/European Union, 2018[23]).While 86% of countries (32 out of 37) have a dedicated validated screening tool for measuring symptoms of depression, and 95% have one for general psychological distress (35 out of 37), only 41% rely on a screening tool for symptoms of anxiety (15 out of 37) (Figure 2.8). Screening tools used by countries vary widely in terms of item length, ranging from two to 40 questions (see Table 2.5).

Variants of the PHQ are the most common screening tool for measuring symptoms of depression, used by 84% (31 out of 37) of countries. The MHI-5 is the most common screening tool for general psychological distress, used by 76% of countries (28 out of 37). In both instances, this is largely driven by Eurostat, which harmonises the data collection efforts of European Union member countries: 26 of the 28 countries that rely on the MHI-5 participate in Eurostat, all but Australia and Israel.[6] The PHQ-8 has been included in

Eurostat's European Health Interview Survey (EHIS), which is conducted every five to six years. Variants of the PHQ are also used by a number of non-European OECD countries (see Table 2.5).

Figure 2.8. Screening tools capturing general psychological distress and symptoms of depression are more commonly used than those for symptoms of anxiety or other disorders

Share of OECD countries that responded to a survey about population mental health and that include measures of risk for mental ill-health in their household surveys, only validated screening tools



Note: Results are shown for all OECD countries except Estonia, which did not participate in the questionnaire. Note that the MHI-5 and PHQ-8 findings are partly driven by Eurostat, although a number of other non-European OECD countries also use these, especially the PHQ-8. The MHI-5 will not be repeated in future EU-SILC ad hoc well-being modules, which will reduce the share of countries regularly collecting it.
Source: Responses to an OECD questionnaire sent to national statistical offices in January 2022.

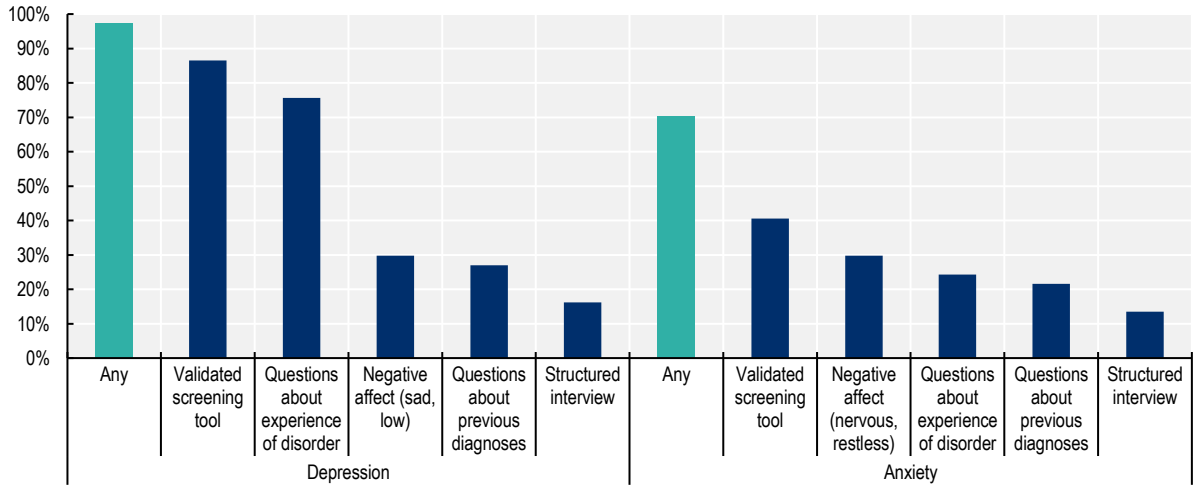
StatLink https://stat.link/exyhv4

OECD countries also collected data on symptoms of anxiety, although often through country-specific tools rather than validated screening tools (Figure 2.9). 70% of countries report capturing anxiety outcomes, through some combination of structured interviews, questions about previous diagnoses or about experience of anxiety disorders, affect data or validated screening tools. Considering all measurement tools included in surveys, more countries indicated using them primarily for measuring symptoms of depression. The only exceptions are questions about negative affect, for which usage is evenly divided: 30% of countries reported using negative affect to measure both anxiety (feeling nervous, anxious) and depression (feeling low, downhearted).

The focus of measurement initiatives on depressive and anxiety disorders reflects the fact that they are some of the most prevalent mental health conditions (OECD/European Union, 2018[23]), and that they contribute highly to the disease burden globally and in OECD countries (Santomauro et al., 2021[24]). Data collection efforts for other specific mental conditions – such as PTSD, bipolar disorder, eating disorders, etc. – remain very uneven across OECD countries (Figure 2.8).

Figure 2.9. Countries do capture anxiety data, but often with non-standardised measures

Share of OECD countries that responded to a survey about population mental health and that include measures of symptoms of depression or anxiety in their household surveys, all tool types



Note: Results are shown for all OECD countries except Estonia, which did not participate in the questionnaire.
Source: Responses to an OECD questionnaire sent to national statistical offices in January 2022.

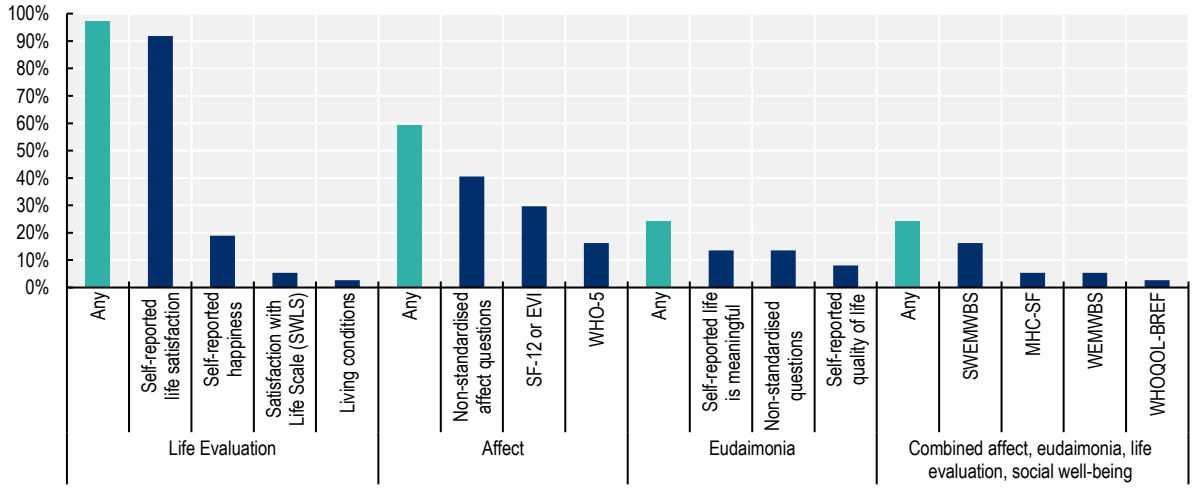
StatLink https://stat.link/6p1esa

Most countries collect comparative data on life evaluation, but less so on affect and eudaimonia

Almost all OECD countries collect some data on life evaluation, primarily through a question on self-reported life satisfaction. Other aspects of positive mental health – affect and eudaimonia – are much less frequently covered by surveys undertaken by OECD countries; even when they are, the tools used are less standardised across countries (Figure 2.10). Measures of affect are more commonly collected than of eudaimonia; 59% of countries collect some form of affect data, through a combination of standardised composite scales and non-harmonised questions, while only 24% collect data on eudaimonia. In terms of standardised tools for measuring positive mental health outcomes, the SF-12 (and the SF-36 sub-component on energy and vitality, EVI), WHO-5 and either WEMWBS or its shorter form SWEMWBS are the three most common instruments; however, their overall use is still low: 30%, 16% and 19% of countries reported using each scale in a household survey, respectively.

Figure 2.10. Affect data are more commonly collected than eudaimonic data, but OECD countries are not aligned in the tools used to collect data on positive mental health beyond life satisfaction

Share of OECD countries that responded to a survey about population mental health and that include measures of positive well-being in their household surveys, all tool types by outcome measure



Note: Results are shown for all OECD countries except Estonia, which did not participate in the questionnaire. Note that the question collected during the EU-SILC 2013 ad hoc well-being module, on the extent to which respondents feel that their life is worthwhile, was not included in this figure given that the question was removed from subsequent well-being modules.
Source: Responses to an OECD questionnaire sent to national statistical offices in January 2022.

StatLink https://stat.link/l8kq9v

There are interesting recent developments in topics such as data collection on mental health awareness

Overall, data collection efforts on additional mental-health related topics (e.g. use of mental health medication and services; mental health of children and young people in the household; loneliness and stress; resilience and self-efficacy; attitudes towards mental health, including stigma and literacy; and questions on unmet needs) are also uneven across countries (see Table 2.11). Many of these issues are not yet well-defined conceptually, with few internationally standardised tools available. For instance, only 30% of countries reported collecting (very different) indicators covering the topics of mental health stigma, discrimination, literacy and knowledge of mental health issues and resources.[7] However, some countries have recently launched new survey efforts – and developed new methods – given increased interest in mental health awareness. For instance, in 2021 Sweden's Public Health Agency conducted an online population survey, covering more than 10 000 respondents, on knowledge and attitudes about mental illness and suicide (Public Health Agency Sweden, 2022[25]). After systematically reviewing more than 400 existing instruments for measuring mental health stigma and conducting cognitive testing, the Public Health Agency concluded that the overwhelmingly negative tone of existing measures was in itself stigmatising and focused mostly on examples of severe mental illness. They hence decided to develop their own survey: the final questionnaire included items that were designed as semantic differentials (word pairs) that captured both positive and negative perceptions of mental illness and focused on all forms of mental illness, including more common experiences of depression, anxiety and stress-related conditions (Public Health Agency Sweden, 2022[25]).

Conclusion and ways forward

Measuring population mental health outcomes is not a new field for producers of official data in OECD countries, and many national statistical offices and health agencies were already collecting relevant data well before COVID-19. Nevertheless, it is also clear that there is room for improvement moving forward.

First, some aspects of mental health are measured more frequently than others, and there is scope for better cross-country harmonisation. The results of the OECD questionnaire to official data producers suggest that existing data collection efforts are not capturing the full range of mental health outcomes – missing aspects of both mental ill-health as well as positive mental health. While 86% of countries use a screening tool for symptoms of depression, and 95% for general psychological distress, only 41% use a standardised screening tool for symptoms of anxiety – and generalised anxiety disorder, along with mood disorders, is one of the most common mental health conditions affecting people in OECD countries. Data collection efforts for other specific mental conditions – such as post-traumatic stress disorder, bipolar disorder, eating disorders, etc. – remain very uneven across countries. When it comes to positive mental health, almost all countries gather some form of life evaluation data, but information about affect and eudaimonia is much less frequently collected (by 59% and 24% of countries, respectively), and often not in a standardised manner. Data producers could hence as a first step expand their use of screening tools to those that include symptoms of anxiety, as well as depression, and move towards more harmonisation for affective and eudaimonic aspects of positive mental health.

Second, it will be important to measure mental health outcomes regularly, and to keep up some of the momentum provided by the high frequency surveys with mental health modules initiated during the first two years of the pandemic. Given the trade-offs between response burden and accuracy that data producers face when choosing between different tools to measure mental health outcomes, adding a single question about people's general mental health status to frequently conducted population surveys could be a way to gather this information regularly and help link data across surveys. Over half of countries (62%) already include such single items in surveys, though question wording varies widely. Canada has been an early leader in developing single-item self-reported mental health (SRMH) indicators, and its question formulation has already been adopted by Chile and Germany, which could make it a useful model for other countries moving forward. While questions about previous diagnoses received by health care professionals are also short, evidence suggests that they focus mostly on people who have been in touch with the health system and hence are better placed in health surveys only.

Chapter 3 reviews the available evidence on the statistical quality of these recommended tools in further detail and provides suggestions for three concrete measures that countries could adapt to maximise international harmonisation and minimise response burden.

Lastly, whichever results are communicated to policy makers or the general public, it is essential to be transparent as to which exact aspect of mental health is being measured, including which areas a specific tool covers and does not cover (e.g. only previous diagnosis? only affect, or also somatic symptoms, and if so, which ones?). This information is important to contextualise findings and to provide transparency as to any limitations that might impact the interpretation of results.

Annex 2.A. Mental health survey measures by country

Table 2.4. Overview of structured interviews to monitor mental health conditions

Focus	Tool	Abbreviation	Number of items	Frame of reference	Time to complete	Already collected by
Diagnosis of mental condition according to ICD-10 and DSM-IV	Composite International Diagnostic Interview	CIDI	More than 300 symptom questions but because of skip rules not all of them are asked to every respondent		75 mins	Australia, Canada, Chile, Germany, Korea, United States (depressive symptoms only)

Note: Results are shown for all OECD countries except Estonia, which did not participate in the questionnaire. For more details on the tool, see Annex 2.B.
Source: Responses to an OECD questionnaire sent to national statistical offices in January 2022.

Table 2.5. Overview of validated screening tools to monitor both general mental ill-health and risk for specific mental health conditions

Focus	Covers	Tool	Abbreviation	Number of items	Frame of reference	Already collected by country
Psychological distress	Negative and positive affect	Mental Health Inventory -5	MHI-5	5	Past month	Australia, **Austria, Belgium, Czech Republic, Denmark,** Finland, France, Germany, **Greece**, Hungary, **Iceland, Ireland,** Israel, Italy, Latvia, **Lithuania, Luxembourg,** Netherlands, **Norway, Poland,** Portugal, **Slovak Republic,** Slovenia, **Spain, Sweden,** Switzerland, **Türkiye, United Kingdom**
Psychological distress	Negative affect, functional impairment	Kessler Scale 10	K10	10	Past 4 weeks	Australia, Canada, Netherlands, New Zealand
Psychological distress	Negative affect	Kessler Scale 6	K6	6	Past 4 weeks	Australia, Japan, Sweden, United States
Psychological distress	Negative and positive affect, somatic symptoms, functional impairment	General Health Questionnaire	GHQ-12	12	Recently	Australia, Belgium, *Finland*, Spain, United Kingdom
Symptoms of depression and anxiety	Negative affect, anhedonia, functional impairment	Patient Health Questionnaire -4	PHQ-4	4 (2 depression, 2 anxiety)	Past 2 weeks	Australia, Belgium, Canada, Chile, Finland, France, Germany, Iceland, Korea, Slovenia, Switzerland, *United Kingdom*, United States

Focus	Covers	Tool	Abbreviation	Number of items	Frame of reference	Already collected by country
Symptoms of depression and anxiety	Negative and positive affect, anhedonia	Hospital Anxiety and Depression Scale	HADS	14 (7 depression, 7 anxiety)	Past week	*France*
Symptoms of depression and anxiety	Negative affect	Hopkins Symptom Checklist	HSCL-5	5	Past week	Norway
Symptoms of depression and anxiety	Negative affect, anhedonia, somatic symptoms, functional impairment	Depression, Anxiety and Stress Scale	DASS-21	21 (7 depression, 7 anxiety, 7 chronic non-specific stress)	Past week	Australia, Italy
Symptoms of depression and anxiety among the general and disabled population	Negative affect, functional impairment	Washington Group on Disability Statistics Short Set on Functioning – Enhanced	WG-SS Enhanced	12 (2 depression, 2 anxiety)	General	Australia, Canada, New Zealand, United States
Symptoms of depression and anxiety among the general and disabled population	Negative affect, functional impairment	Washington Group Extended Set on Functioning	WG-ES	37 (3 depression, 3 anxiety)	General	United States
Depressive symptoms	Negative affect, anhedonia, somatic symptoms, functional impairment (matched to major depressive disorder per DSM-IV and DSM-5 criteria)	Patient Health Questionnaire -8	PHQ-8	8	Past 2 weeks	Austria, **Czech Republic, Denmark,** Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, **Netherlands**, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, **Sweden**, Türkiye, United Kingdom, United States
Depressive symptoms	Negative affect, anhedonia, somatic symptoms, functional impairment (matched to major depressive disorder per DSM-IV and DSM-5 criteria)	Patient Health Questionnaire -9	PHQ-9	9 (PHQ-8 + question on suicidal ideation)	Past 2 weeks	Australia, Belgium, Canada, Finland, France, Germany, Italy, Korea, Slovenia, Switzerland, United States
Depressive symptoms	Negative affect, anhedonia	Patient Health Questionnaire -2	PHQ-2	2	Past 2 weeks	Australia, Canada, Chile, Finland, Germany, Italy, Norway, United States
Depressive symptoms	Negative and positive affect, anhedonia, somatic symptoms, functional impairment, interpersonal challenges	Center for Epidemiological Studies Depression Scale	CES-D	20	Past week	Mexico
Symptoms depression among recent mothers	Negative and positive affect, anhedonia, functional impairment	Edinburg Post-natal Depression Scale	EPDS	6	Past week	Italy
Symptoms of anxiety	Negative affect, somatic symptoms, functional impairment	Generalised Anxiety Disorder-7	GAD-7	7	Past 2 weeks	Australia, Belgium, Canada, Finland, France, Germany, Iceland, Korea, Slovenia,

Focus	Covers	Tool	Abbreviation	Number of items	Frame of reference	Already collected by country
						Switzerland, United States
Symptoms of anxiety	Negative affect, functional impairment	Generalised Anxiety Disorder-2	GAD-2	2	Past 2 weeks	Australia, Canada, Chile, Germany, Mexico, *United Kingdom*, United States
Symptoms of anxiety	Negative affect,(including panic-like anxiety), functional impairment, subjective well-being	The State and Trait Anxiety Scale	STAI	40 (20 state anxiety, 20 trait anxiety)	State anxiety: "in this moment", trait anxiety: "generally"	Italy
Symptoms of panic disorder	Presence and severity of anxiety attacks, somatic symptoms	Patient Health Questionnaire-Panic Disorder	PHQ-PD	15	Past 4 weeks	Germany, Switzerland
Symptoms of post-traumatic stress disorder (PTSD)	Presence and severity of PTSD symptoms (matched to DSM-5 criteria)	PTSD Checklist for DSM-5	PCL-5	20	Past 4 weeks	Canada
Symptoms of PTSD	Presence and severity of PTSD symptoms (matched to DSM-5 criteria)	Primary Care PTSD Screen for DSM-5	PC-PTSD-5	5	Past 4 weeks	Switzerland
Symptoms of PTSD	Presence and severity of PTSD symptoms (matched to DSM-IV criteria)	Impact of Event Scale – revised	IES-R	22	Past week	Italy
Symptoms of agoraphobia	Presence and severity of anxiety related to different aspects of everyday life	Angstbarometer	Angstbarometer	12	Past year	Switzerland
Symptoms of social anxiety disorder	Presence and severity of symptoms of social anxiety disorder	Mini-Social Phobia Inventory	Mini-SPIN	3	Past week	Finland, Switzerland
Symptoms of substance abuse or addiction	Presence and severity of symptoms of alcoholism	CAGE Substance Abuse Screening Tool	CAGE	4	No specific recall period	Belgium
Symptoms of substance abuse or addiction	Presence and severity of symptoms of alcoholism	Alcohol Use Disorders Identification Test-Concise	AUDIT-C	3	No specific recall period	Chile, Sweden
Symptoms of substance abuse or addiction	Presence and severity of symptoms of alcoholism	Alcohol Use Disorders Identification Test	AUDIT	10	No specific recall period	France, Spain
Symptoms of substance abuse or addiction	Presence and severity of Internet addiction and compulsive, pathological, or problematic online behaviours (matched to DSM-IV criteria for substance addiction and pathological gambling)	Compulsive Internet Use Scale	CIUS	14	No specific recall period	Switzerland
Symptoms of eating disorders	Presence and severity of symptoms of anorexia nervosa and	SCOFF	SCOFF	5	Past 3 months	Belgium, Finland, Germany

Focus	Covers	Tool	Abbreviation	Number of items	Frame of reference	Already collected by country
	bulimia nervosa					
Symptoms of eating disorders	Presence and severity of symptoms of binge eating disorder, bulimia nervosa and recurrent binge eating	Patient Health Questionnaire-Eating Disorder Module	PHQ-ED	6	Past 3 months	France
Symptoms of eating disorders in 7-17 year-olds	Presence and severity of symptoms of eating disorders	Screening questions from the Development and Wellbeing Assessment – Eating Disorder Module	DAWBA	5	No specific recall period	United Kingdom

Note: Countries in italics are those that have explicitly stated that they no longer collect the measure in question. Countries in bold did not report collecting the instrument in their official questionnaire submission, however, it was added by the OECD Secretariat based on the country's participation in the European Health Interview Survey (EHIS), which contains the PHQ-8 as a core module. The PHQ-4 country practice was added in by the Secretariate for countries collecting both the PHQ and GAD (from which the PHQ-4 pulls its indicators), regardless of individual country reporting on the PHQ-4 itself. Results are shown for all OECD countries except Estonia, which did not participate in the questionnaire. Data for the United Kingdom include only surveys carried out by the Office for National Statistics on mental health and do not include the data collected by devolved administrations. For details of the tools collected by at least two OECD countries, see Annex 2.B.
Source: Responses to an OECD questionnaire sent to national statistical offices in January 2022.

Table 2.6. Overview of questions about previous diagnoses

Category	Example question framing	Answer options	Frame of reference	Already collected by country
Received diagnosis of any mental health condition	Have you been told by a doctor or nurse that you have any of these long-term health conditions? List: Mental health condition (including depression or anxiety) (AUS) Have your mental health problems ever been diagnosed as a mental disorder by a professional (psychiatrist, doctor, clinical psychologist)? (SVN)	Yes / No	Lifetime	Australia, Slovenia
Received diagnosis of any mood disorder (including depression)	Have you ever in your life been diagnosed by a doctor with any of the following health problems or illnesses? In the event that you have been diagnosed any of them, have you received or are you undergoing medical treatment? Depression or anxiety (CHL) During your life, has a doctor ever told you that you had a psychiatric or psychological disorder or an addiction? Depression or depressive episode (FRA)	Yes / No	Lifetime, last 12 months, or during COVID	Australia, Austria, Canada, Chile, Costa Rica, France, New Zealand, Slovenia, Spain, United States
Received diagnosis of anxiety disorder	Have you ever been told by a doctor, nurse or other health professional that you have any of these conditions? Anxiety (AUS) Has a health professional ever told you that you have…? Chronic anxiety (CRI) During your life, has a doctor ever told you that you had a psychiatric or psychological disorder or an addiction? Anxiety disorder (generalised anxiety, phobia, obsessive	Yes / No	Lifetime, during COVID	Australia, Chile, Costa Rica, France, New Zealand, Slovenia, Spain, United States

Category	Example question framing	Answer options	Frame of reference	Already collected by country
	compulsive disorder, etc.) (FRA)			
Received diagnosis of bipolar disorder or mania	Have you ever been told by a doctor that you have bipolar disorder, which is sometimes called manic depression? (NZL)	Yes / No	Lifetime, during COVID	Australia, France, New Zealand, Slovenia
Received diagnosis of post-traumatic stress disorder (PTSD)	Have you ever been diagnosed with PTSD? (CAN)	Yes / No	Lifetime	Australia, Canada
Received diagnosis of obsessive compulsive disorder (OCD)	Have your mental health problems ever been diagnosed as a mental disorder by a professional (psychiatrist, doctor, clinical psychologist)? Obsessive compulsive disorder (SVN)	Yes / No	Lifetime	Australia, Slovenia
Received diagnosis of schizophrenia or other psychotic disorders	During your life, has a doctor ever told you that you had a psychiatric or psychological disorder or an addiction? Schizophrenia (FRA)	Yes / No	Lifetime, during COVID	France, Slovenia
Received diagnosis of personality disorder	During your life, has a doctor ever told you that you had a psychiatric or psychological disorder or an addiction? borderline personality disorder (FRA)	Yes / No	Lifetime, during COVID	France
Received diagnosis of agoraphobia or social disorder	Were you told by a doctor, nurse or other health professional that you had [...] mental health condition? Agoraphobia (AUS)	Yes / No	Lifetime	Australia
Received diagnosis of addictive disorder or substance abuse problems	Have you ever been told by a doctor, nurse or other health professional that you have any of these conditions? Harmful use or dependence on alcohol or drugs (AUS) During your life, has a doctor ever told you that you had a psychiatric or psychological disorder or an addiction? Addiction or addictive disorder (FRA)	Yes / No	Lifetime, during COVID	Australia, France
Received diagnosis of an eating disorder	Have your mental health problems ever been diagnosed as a mental disorder by a professional (psychiatrist, doctor, clinical psychologist)? Eating disorder (SVN)	Yes / No	Lifetime, during COVID	France, Slovenia
Received diagnosis of conduct disorder or behavioural / emotional problems	Have you ever been told by a doctor, nurse or other health professional that you have any of these conditions? Behavioural or emotional problems (AUS) Have you ever been diagnosed with conduct disorders by a medical professional? (ESP)	Yes / No	Lifetime	Australia, Spain
Neurodiversity: received diagnosis of attention deficit hyperactivity disorder (ADHD)	Have [you/name] ever been told by a doctor or other health professional that {you/he/she} had attention deficit hyperactivity disorder (ADHD) or attention deficit disorder (ADD)? (USA)	Yes / No	Lifetime	Germany, United States
Neurodiversity: received diagnosis of	Have you ever been diagnosed with autism by a medical professional? (ESP)	Yes / No	Lifetime	Spain

Category	Example question framing	Answer options	Frame of reference	Already collected by country
autism spectrum disorder (ASD)				
Received diagnosis of any other mental health condition	Do you have any other long-term physical or mental health condition that has been diagnosed by a health professional? (CAN) Have your mental health problems ever been diagnosed as a mental disorder by a professional (psychiatrist, doctor, clinical psychologist)? (SVN)	Yes / No	Lifetime, last 12 months, or during COVID	Australia, Canada, Chile, Costa Rica, France, Slovenia

Note: Results are shown for all OECD countries except Estonia, which did not participate in the questionnaire. Data for the United Kingdom include only surveys carried out by the Office for National Statistics on mental health and do not include the data collected by devolved administrations.
Source: Responses to an OECD questionnaire sent to national statistical offices in January 2022.

Table 2.7. Overview of questions about experienced symptoms and mental health conditions

Category	Example question framing	Answer options	Frame of reference	Already collected by country
Self-reported mental health problems	Have you suffered from psychological stress or an acute illness in the last three months? (ISR) Are you currently facing mental health problems? (SVN) Do you have your own experience with mental illness? (SWE) Do you think you ever had a problem with your own mental health? (USA)	Yes / No	Lifetime, last 12 months, last 3 months	Hungary, Israel, Slovenia, Sweden, United States
Self-reported mood disorder (depression, etc.) or mood disorder symptoms	During the past 12 months, have you had any of the following diseases or conditions? Depression (European OECD countries participating in EHIS) Do you have a mood disorder? (CAN) Next I will ask you some questions related to different chronic diseases or health conditions that you may currently have. Chronic diseases are those of long duration and usually evolve slowly.	Yes / No	Lifetime, last 12 months, current	Australia, Austria, **Belgium**, Canada, Costa Rica, **Czech Republic, Denmark, France,** Germany, Greece, Hungary, **Iceland**, Ireland, **Italy, Latvia, Lithuania, Luxembourg,** Netherlands, Norway**, Poland, Portugal, Slovak Republic,** Slovenia, Spain, Sweden, Türkiye, **United Kingdom,** United States

Category	Example question framing	Answer options	Frame of reference	Already collected by country
	Do you have chronic depression? (CRI)			
Self-reported anxiety disorder, or anxiety symptoms	Do you have an anxiety disorder? (CAN) During the last 12 months did you have or do you have any of the chronic diseases / diseases that are listed: Anxiety disorders (e.g. panic attacks, anxiety) (GRC) Have you ever suffered from chronic anxiety? (ESP)	Yes / No	Lifetime, last 12 months, last 3 months, current	Australia, Canada, Costa Rica, Greece, Hungary, Norway, Slovenia, Spain, Sweden
Self-reported bipolar disorder or mania	Do you have any of these conditions? Bipolar disorder (AUS) Do you have a mood disorder such as depression, bipolar disorder, mania or dysthymia? (CAN)	Yes / No	Lifetime	Australia, Canada
Self-reported PTSD	Do you currently experience symptoms of PTSD? (CAN)	Yes / No	Lifetime	Australia, Canada
Self-reported OCD	Do you have any of these conditions? Obsessive-compulsive disorder (OCD) (AUS)	Yes / No	Lifetime	Australia
Self-reported schizophrenia or other psychotic disorders	(Apart from any conditions you have told me about) do you have any other mental health, behavioural or cognitive conditions, such as these? Schizophrenia (AUS) [Do you have] Schizophrenia, schizotypal and delusional disorders (HUN)	Yes / No	Lifetime or last 12 months	Australia, Hungary
Self-reported agoraphobia or social disorder	Do you have any of these conditions? Agoraphobia (AUS)	Yes / No	Lifetime	Australia
Self-reported addictive disorder or substance abuse problems	(Apart from any conditions you have told me about) do you have any other mental health, behavioural	Yes / No	Lifetime or last 12 months	Australia, Hungary

Category	Example question framing	Answer options	Frame of reference	Already collected by country
	or cognitive conditions, such as these? Dependence on alcohol; Dependence on drugs; Harmful use or dependence on medicinal, prescription drugs (AUS)			
Self-reported eating disorder	In the past 12 months, how often have you done the following things? a. Been preoccupied with a desire to be thinner b. Vomited to lose weight c. Changed your eating habits in order to manage your weight (CAN)	Never / A few times / Monthly / Weekly / Daily	Last 12 months	Canada
Self-reported conduct disorder or behavioural / emotional problems	Have you suffered from conduct disorders in the last 12 months? (ESP)	Yes / No	Lifetime or last 12 months	Australia, Spain
Self-reported ADHD	(Apart from any conditions you have told me about) do you have any other mental health, behavioural or cognitive conditions, such as these? Attention Deficit Hyperactivity Disorder (ADHD) (AUS)	Yes / No	Lifetime	Australia
Self-reported ASD	Have you suffered from autism in the last 12 months? (ESP)	Yes / No	Lifetime or last 12 months	Australia, Spain
Self-reported dementia	(Apart from any conditions you have told me about) do you have any other mental health, behavioural or cognitive conditions, such as these? Dementia, including Alzheimer's Disease (AUS)	Yes / No	Lifetime	Australia
Self-reported intellectual impairment	(Apart from any conditions you have told me about) do you have any other mental health, behavioural or cognitive conditions, such as these? Intellectual impairment, mental retardation (AUS)	Yes / No	Lifetime	Australia
Self-reported learning disorder	(Apart from any conditions you have told me about) do you have any other mental health, behavioural or cognitive conditions, such as these? Learning	Yes / No	Lifetime	Australia

Category	Example question framing	Answer options	Frame of reference	Already collected by country
	difficulties, including dyslexia (AUS)			
Self-reported other mental disorder	(Apart from any conditions you have told me about) do you have any other mental health, behavioural or cognitive conditions, such as these? Any other mental or behavioural condition (AUS)	Yes / No	Lifetime or last 12 months	Australia, Costa Rica, Hungary

Note: Results are shown for all OECD countries except Estonia, which did not participate in the questionnaire. Data for the United Kingdom include only surveys carried out by the Office for National Statistics on mental health and do not include the data collected by devolved administrations.
Source: Responses to an OECD questionnaire sent to national statistical offices in January 2022.

Table 2.8. Overview of questions about suicidal ideation and suicide attempts

Category	Example question framing	Answer options	Frame of reference	Already collected by country
Suicidal ideation	Final question of the PHQ-9	Not at all / Several days / More than half the days / Nearly every day	Last 2 weeks	Australia, Belgium, Canada, Finland, France, Germany, Italy, Korea, Slovenia, Switzerland, United States
Suicidal ideation	Have you seriously contemplated suicide since the COVID-19 pandemic began? (CAN) Have you had this experience [seriously considering suicide] in the last 12 months? (CHL) In the last 12 months, have you thought about committing suicide? (FRA) Have you ever been in a situation where you seriously considered taking your own life? (SWE)	Yes / No	Lifetime, last 12 months, during COVID	Australia, Belgium, Canada, Chile, Finland, France, Korea, Mexico, Slovenia, Sweden, Switzerland, United States
Self-harm behaviours	Sometimes people harm themselves on purpose but they do not mean to take their life. In the past 12 months, did you ever harm yourself on purpose but not mean to take your life? (CAN)	Yes / No	Lifetime, last 12 months, last 2 weeks	Australia, Canada, Finland, Greece, Mexico

Category	Example question framing	Answer options	Frame of reference	Already collected by country
	During the past 2 weeks, how often did you have thoughts of hurting yourself? (GRC)	Not at all/ Several days/ More than half the days/ Nearly every day		
Suicide attempts	Did you attempt to commit suicide in the last 12 months? (BEL) Have you attempted to actually commit suicide over the last 12 months? (KOR) Have you ever attempted suicide? (LUX)	Yes / No	Lifetime, last 12 months, during COVID	Australia, Belgium, Canada, Chile, Finland, France, Korea, Luxembourg, Sweden, United States
Suicide attempt led to hospitalisation or required medical care	Did you stay in a hospital overnight or longer because you tried to kill yourself? (USA)	Yes / No	Lifetime or last 12 months	Australia, United States
Received counselling following suicidal thoughts or suicide attempt	Following your thoughts of suicide, did you talk to anyone? (CHE) During the past 12 months, did you get medical attention from a doctor or other health professional as a result of an attempt to kill yourself? (USA)	Yes / No	Lifetime or last 12 months	France, Switzerland, United States

Note: Results are shown for all OECD countries except Estonia, which did not participate in the questionnaire. Data for the United Kingdom include only surveys carried out by the Office for National Statistics on mental health and do not include the data collected by devolved administrations.
Source: Responses to an OECD questionnaire sent to national statistical offices in January 2022.

Table 2.9. Overview of questions about general mental health status

Category	Example question framing	Answer options	Frame of reference	Already collected by country
Self-reported general mental health status	In general, how is your mental health? (CAN, CHL, DEU) How is your mental state, usually? (ISR)	Excellent / Very good / Good / Fair / Poor (AUS, CAN, CHL, DEU) Very good / good / not so good / Not good at all (ISR)	Current or last 4 weeks	Australia, Canada, Chile, Costa Rica, Finland, Germany, Iceland, Israel, Slovenia, Switzerland, United States
Self-reported number of mentally healthy days	During the past 30 days, how often was your mental health not good? (USA)	[Number of days]	Last 30 days	United States
Self-reported recovery	At this time do you consider yourself to be in recovery or recovered from your own mental health problem?	Yes / No	General assessment	United States

Category	Example question framing	Answer options	Frame of reference	Already collected by country
	(USA)			
Self-reported satisfaction with mental health status	On a scale from 1 to 10 can you indicate to what extent you are satisfied with your mental health? (NLD) How satisfied are you with your mental health? (NOR)	0 (completely dissatisfied) to 10 (completely satisfied)	General assessment	Netherlands, Norway
Self-reported mental health status and COVID-19	Compared to before the pandemic started, how would you say your mental health is now? (CAN) Has your mental health/well-being been affected by the COVID-19 pandemic during 2020 / during the last 12 months? (DNK, LVA, PRT, SVK, SVN, TUR) How has your morale been affected by the pandemic? (CHE) During the periods of confinement, have there been times when you have felt so discouraged that nothing could cheer you up? (FRA)	Much better now / Somewhat better now / About the same / Somewhat worse now / Much worse now (CAN) 1. Yes, has been negatively affected 2. Yes, has been positively affected 3. No, has not been affected (DNK, LVA, PRT, SVK, TUR) 0 (much worse) - 10 (much better) (CHE) Yes / No (FRA)	During COVID-19	Canada, Denmark, Finland, France, Germany, Israel, Japan, Latvia, Netherlands, Portugal, Slovak Republic, Slovenia, Switzerland, Türkiye
Mental health interferes with daily activities (impairment-days)	Does your mental state interfere with your daily life at work? With family? (ISR) Have you felt very sad or hopeless for more than two weeks over the last 12 months to a degree that you have experienced disruptions in your daily life? (KOR) During the past 12 months, did you ever feel so sad or hopeless almost every day for two weeks or more in a row that you stopped doing some usual activities? (USA)	Yes / No	Varies from past 12 months to past 4 weeks	Australia, Canada, Hungary, Israel, Korea, Spain, United States

Note: Results are shown for all OECD countries except Estonia, which did not participate in the questionnaire. Data for the United Kingdom include only surveys carried out by the Office for National Statistics on mental health and do not include the data collected by devolved administrations.
Source: Responses to an OECD questionnaire sent to national statistical offices in January 2022.

Table 2.10. Overview of indicators of positive mental health

Components	Tool	Abbreviation	Number of items	Frame of reference	Already collected by country
Positive affect	WHO-5 Wellbeing Index	WHO-5	5	Last 2 weeks	France, Hungary, Italy, Latvia, New Zealand, Slovenia
Positive and negative affect, functional impairment (Mental Health Component Summary)	Short Form Health Status	SF-12	12	Last 4 weeks	Chile, Italy, Netherlands, New Zealand, *Spain*, United States
Positive and negative affect, functional impairment	SF-36	SF-36	36	Last 4 weeks	Australia, Germany
Positive and negative affect	SF-36 vitality sub-scale	EVI	4	Last 4 weeks	Australia, Belgium, Italy, Switzerland
Positive or negative affect	Non-standardised affect questions Example questions: During the day yesterday, did you feel happy? (FRA) During this period [last 12 months], to what extent did you experience the following feelings? Stress and anxiety (ISR) Now, I am going to mention a series of emotions or feelings. How often have you felt… during the last two weeks? Angry Optimistic Worried Happy Sad Calm Tired Useful (CHL)	NA 0 (least happy) – 10 (happiest) (FRA) To a large extent / Certain / Not so much / Not at all (ISR) Never, almost never sometimes, almost always or always (CHL)	Varies from 1 to 8	Varies from yesterday to last year	Chile, Costa Rica, Finland, France, Ireland, Israel, Italy, Japan, Latvia, Netherlands, New Zealand, Norway, Slovenia, Sweden, United Kingdom
Eudaimonia	Self-reported feeling that life is worthwhile or meaningful Example questions: Do you feel that what you do in your life has meaning, value? Answer on a scale of 0 (no meaning) to 10 (full of meaning) (FRA)	NA	1	General assessment	Australia, **Austria, Belgium,** Canada, **Czech Republic, Denmark, Finland**, France, **Germany, Greece,** Hungary, **Iceland, Ireland, Italy, *Japan*, Latvia, Lithuania, Luxembourg, Netherlands, *New Zealand*, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Türkiye,** United Kingdom

Components	Tool	Abbreviation	Number of items	Frame of reference	Already collected by country
	How would you usually describe yourself? Would you say: 1: Happy and interested in life 2: Somewhat happy 3: Somewhat unhappy 4: Unhappy with little interest in life 5: So unhappy that life is not worthwhile (CAN)				
Eudaimonia	Self-reported quality of life Example question: Would you rate your quality of life as... ? 1: Excellent 2: Very good 3: Good 4: Fair 5: Poor (CAN)	NA	1	General assessment	Canada, Costa Rica, Finland, Switzerland
Eudaimonia	Self-reported satisfaction with self Example question: How satisfied are you with..... yourself? (CRI)	1 Very satisfied, 2 Satisfied, 3 Moderately satisfied, 4 Dissatisfied, 5 Very dissatisfied	1	General assessment	Costa Rica, Finland
Eudaimonia	Self-reported sense of purpose, accomplishment or achievement of goals Example questions: So far, I have achieved the goals that are important to me in life (MEX) My life has a clear sense of purpose (USA) Most days I feel a sense of accomplishment from what I do (USA)	NA	1	General assessment	***Mexico, United States***
Eudaimonia	Self-reported sense of being a beneficial participant of society	NA	1	General assessment	***Hungary***

Components	Tool	Abbreviation	Number of items	Frame of reference	Already collected by country
	Example question: How do you feel about yourself being an important and beneficial participant of the society? (HUN)				
Life evaluation	Self-reported life satisfaction Example question: Overall, how satisfied are you with life as a whole these days? Please answer on a scale from 0 to 10. 0 means "not at all satisfied" and 10 means "completely satisfied". (European OECD countries participating in EU-SILC well-being modules)	NA	1	General assessment	***Australia,*** **Austria,** Belgium, Canada, Chile, **Czech Republic, Denmark,** Finland, France, **Germany, Greece,** Hungary, **Iceland,** Ireland, Israel, Italy, ***Korea***, Latvia, **Lithuania, Luxembourg,** ***Mexico*****,** Netherlands, **New Zealand, Norway, Poland, Portugal, Slovak Republic,** Slovenia, **Spain, Sweden,** Switzerland, **Türkiye,** United Kingdom, ***United States***
Life evaluation	Satisfaction with Life Scale (SWLS)	SWLS	5	General assessment	Norway, Slovenia
Life evaluation	Self-reported happiness Example question: Overall, how happy do you think you are? Please check one box on a scale of 1- 10 where 1 means very unhappy and 10 very happy. (ISL)	NA	1	General assessment	Chile, France, Iceland, ***Japan***, Netherlands, Switzerland, ***United States***
Life evaluation	Self-reported living conditions Example question: Currently the living conditions in your household are: 1. Very good; 2. Good; 3. Fair; 4. Bad (COL)	NA	1	General assessment	***Colombia***
Positive affect, eudaimonia, life satisfaction, social well-being	Mental Health Continuum Short Form	MHC-SF	14	Past month	Canada, Slovenia
Positive affect, eudaimonia, social well-being	Warwick- Edinburgh Mental Well-Being Scale	WEMWBS	14	Last 2 weeks	Finland, Norway*

Components	Tool	Abbreviation	Number of items	Frame of reference	Already collected by country
Positive affect, eudaimonia, social well-being	Short Warwick- Edinburgh Mental Well-Being Scale	SWEMWBS	7	Last 2 weeks	Canada, Finland, Germany, Iceland, Sweden, United Kingdom
Positive and negative affect, eudaimonia, self-esteem, concentration	WHO Quality of Life-BREF psychological health domain	WHOQOL-BREF	6	Last 2 weeks	Chile

Note: *Norway does not currently collect WEMWBS but indicated that the tool may be included in future rounds of the National Survey on Quality of Life. Countries in italics are those that have explicitly stated that they no longer collect the measure in question. Countries in bold did not report collecting the instrument in their official questionnaire submission, however, it was added by the OECD Secretariat based on the country's participation in the European Union Statistics on Income and Living Conditions (EU-SILC), which contained the question "Overall, to what extent do you feel that the things you do in your life are worthwhile?" in the 2013 ad-hoc module focusing on well-being; the measure was not included again in 2018. Countries in bold and italics did not report collecting the instrument in their official questionnaire, however, it was added by the OECD Secretariat based on the country's participation in a 2016 OECD questionnaire on subjective well-being measures. Results are shown for all OECD countries except Estonia, which did not participate in the questionnaire. Data for the United Kingdom include only surveys carried out by the Office for National Statistics on mental health and do not include the data collected by devolved administrations. For details of the tools collected by at least two OECD countries, see Annex 2.B.
Source: Responses to an OECD questionnaire sent to national statistical offices in January 2022.

Table 2.11. High-level overview of additional mental health-related topics collected by countries

Topic Area	Types of Tools Used	Types of Indicators Collected	Already collected by country
Access to / use of mental health services	Self-reported non-standardised questions	Sought care from a mental health professional (psychologist, psychiatrist, etc.)	Australia, Canada, Chile, Colombia, Finland, France, Ireland, Japan, Korea, Luxembourg, New Zealand, Slovenia, United States
		Medication prescribed or taken (anti-depressants, anxiolytics)	Belgium, Canada, Chile, Finland, France, Germany, Norway, Slovenia, Spain
Mental health of children and young people	Standardised screening tools, diagnoses and experienced symptoms	Strengths and Difficulties Questionnaire (SDQ); KIDSCREEN-27 and KIDSCREEN-10; Screen for Child Anxiety and Related Emotional Disorders (SCARED); Short Moods and Feelings Questionnaire (SMFQ)	Australia, Belgium, Finland France, Germany, Italy, Slovenia, Spain, United Kingdom, United States
		Diagnostic and reported experience of conduct disorders, behavioural and emotional issues, positive and adverse early childhood experiences, and substance use/abuse behaviours	Canada, Italy, Spain, Türkiye, United States
Loneliness and stress	Standardised screening tools, non-standardised self-reported indicators	Loneliness and social connections: UCLA Loneliness Scale, Oslo Social Support Scale; Multidimensional Scale of Perceived Social Support; non-standardised indicators	Australia, **Austria, Belgium,** Canada, Colombia, Costa Rica, **Czech Republic, Denmark,** Finland, France, Germany, **Greece,** Hungary, Iceland, Ireland, Israel, Italy, Japan, Latvia, **Lithuania, Luxembourg,** Netherlands, New Zealand, **Norway, Poland, Portugal,** Slovak Republic, Slovenia, **Spain,** Sweden, United States, **United Kingdom**
		Stress: Cohen Perceived Stress Scale (PSS); non-standardised indicators	Canada, Colombia, Iceland, Israel, Italy, Korea, Latvia, Slovenia, Sweden
Resilience, optimism and self-efficacy	Standardised composite scales, non-standardised self-reported indicators	Pearlin and Schooler's Mastery Scale, General Self-Efficacy Scale, Brief Resilient Coping Scale, Short Sense of Coherence Questionnaire, Connor-Davidson Resilience Scale (CD RISC-10), Single Item Self-esteem Scale; non-standardised indicators	Australia, Canada, Germany, Italy, Norway, Slovenia, Switzerland

Attitudes towards mental health	Standardised composite scales, non-standardised self-reported indicators	Non-standardised indicators covering topics of stigma, discrimination, literacy and knowledge of mental health issues and resources	Costa Rica, Hungary, Italy, Japan, Korea, Mexico, New Zealand, Norway, Slovenia, Sweden
		Mental health literacy: Depression and Anxiety Literacy questionnaire (D-Lit; A-Lit)	Slovenia

Note: Results are shown for all OECD countries except Estonia, which did not participate in the questionnaire. Data for the United Kingdom include only surveys carried out by the Office for National Statistics on mental health and do not include the data collected by devolved administrations. Countries in bold did not report collecting the instrument in their official questionnaire submission, however, it was added by the OECD Secretariat based on the country's participation in the European Health Interview Survey (EHIS), which contained the Oslo Social Support Scale (OSS-3) in waves 2 and 3.
Source: Responses to an OECD questionnaire sent to national statistical offices in January 2022.

Annex 2.B. Details on standardised survey tools to measure mental health

Mental ill-health

Mental health conditions: Structured interviews

Composite International Diagnostic Interview (CIDI): The Composite International Diagnostic Interview (CIDI) is a comprehensive, fully-structured interview designed to be used by trained lay interviewers for the assessment of mental disorders according to the definitions and criteria of ICD-10 and DSM-IV (Kessler and Bedirhan Üstün, 2006[26]). A computer-assisted version of the interview is available along with a direct data entry software system that can be used to keypunch responses to the paper-and-pencil version of the interview. The CIDI is intended for use in epidemiological and cross-cultural studies as well as for clinical and research purposes. It allows investigators to measure the prevalence of lifetime and 12-month mental conditions, the severity and courses of these disorders, their impact on home management, work life, relationships and social life, and service and medications use. Several versions of the CIDI exist, but the latest version is the World Health Organization's Composite International Diagnostic Interview (WHO-CIDI) V3.0 (Harvard Medical School, n.d.[27]). In total, the CIDI consists of a screening module and 40 sections, 22 of which are diagnostic sections to assess mood (two sections), anxiety (seven sections), substance abuse (two sections), childhood (four sections) and other disorders (seven sections). The remaining sections assess functioning and physical comorbidity, risk factors, socio-demographic information and the treatment of mental disorders. The screening module, which includes a series of introductory questions about the respondent's general health before delving into the diagnostic stem questions, has been shown to increase the accuracy of diagnostic assessments by reducing the effects of respondent fatigue and unwillingness to disclose on stem question endorsement (Harvard Medical School, n.d.[27]).

Symptoms of mental ill-health: Screening tools

The public health tools presented in this section focus mainly on royalty-free instruments, since fees and copyright restrictions might present a barrier to use.

Mental Health Inventory (MHI-5): The Mental Health Inventory-5 (MHI-5) is a five-item scale to screen for symptoms of psychological distress. It is drawn from the 38-item Mental Health Inventory (MHI) and included in the 20-item and 36-item versions of the Short Form Health Survey (SF-20 and SF-36) (Berwick et al., 1991[28]; Kelly et al., 2008[29]). The questions tap into both negative and positive affect, with three items focusing on low/depressed mood and two on nervousness/anxiety (although the tool itself is not used to present these aspects separately). The MHI-5 has been found to be a reliable measure of mental health status and has been validated against both depressive and, to a lesser degree, also anxiety disorders (including generalised anxiety and panic disorder) in general population and patient samples in a range of countries (Yamazaki, Fukuhara and Green, 2005[30]; Hoeymans et al., 2004[31]; Elovanio et al., 2020[32]; Gill et al., 2007[33]; Rumpf et al., 2001[34]; Strand et al., 2003[35]; Thorsen et al., 2013[36]). There is some evidence that removing the two anxiety-related items does not reduce the effectiveness of the MHI in detecting depression, although this has not been examined in studies in which a formal diagnosis according to clinical criteria was used as a gold standard (Yamazaki, Fukuhara and Green, 2005[30]).

Table 2.12. MHI-5 Questionnaire with scoring breakdown

	All of the time	Most of the time	A good bit of the time	Some of the time	A little of the time	None of the time
During the past month, how much of the time:						
1. Have you been a happy person? (reverse coded)	1	2	3	4	5	6
2. Have you felt calm and peaceful? (reverse coded)	1	2	3	4	5	6
3. Have you been a very nervous person?	1	2	3	4	5	6
4. Have you felt downhearted and blue?	1	2	3	4	5	6
5. Have you felt so down in the dumps that nothing could cheer you up?	1	2	3	4	5	6

Note: All items are added together to provide a total score from 5 to 30, which is then transformed into a variable ranging from 0-100 using a standard linear transformation. Higher values indicate better mental health, with the following cut-off points for various degrees of psychological distress: 68 or less mild, moderate or severe, 60 or less moderate or severe, 52 or less severe.
Source: Kelly, M.J. et al. (2008[29]), "Evaluating cutpoints for the MHI-5 and MCS using the GHQ-12: A comparison of five different methods", *BMC Psychiatry* Vol. 8/10, https://doi.org/10.1186/1471-244X-8-10.

***The Short-Form Health Survey (SF-12)**:* The Short-Form Health Survey (SF-12) is a tool to measure health-related quality of life. It was developed as a shorter alternative to the SF-36 questionnaire to be used in the general population and in large surveys and contains up to two items for each of the SF-36's eight dimensions: general mental health, energy and fatigue, bodily pain, general health perceptions, limitations on physical activity due to health, limitations on social activity due to physical or emotional conditions, limitations on day-to-day activities due to physical health, and limitations on day-to-day activities due to emotional health (Ware et al., 2002[37]). A number of questions in both the SF-12 and SF-36 are taken directly from the Mental Health Inventory (MHI), which also features the MHI-5 free-standing scale in its own right (see above) (RAND, n.d.[38]). Two summary scores, the Physical Component Summary (PCS) and the Mental Component Summary (MCS), can be derived from the SF-12, and a range of scoring methods have been validated against both active and recent depressive disorders and to a lesser degree also anxiety disorders in general population samples (Ware et al., 2002[37]; Gill et al., 2007[33]; Vilagut et al., 2013[39]). Some evidence suggests that the association between the SF-12's physical health dimensions might be more strongly related with mental health in low-income settings, with implications for context-specific weights (Ohrnberger et al., 2020[40]). The SF-12 is subject to copyright restrictions and can thus not be republished in this report (Quality Metric, n.d.[41]).

Kessler Scale (K10/ K6): The Kessler psychological distress scale, which is most often used in its 10-item (K10) and 6-item (K6) form, is a screening tool for identifying adults with significant levels of psychological distress. The questions focus on somatic symptoms and negative affect, particularly on both low-depressed mood and nervousness/anxiety. While these aspects are usually not presented separately and a total score for distress is usually used, factor analysis has established depression and anxiety as distinct clusters in the K10 (Brooks, Beard and Steel, 2006[42]). Indeed, although it is often applied in primary clinical settings as well, it was designed for use in the general population, and sensitivity and specificity analysis support both K6 and K10 as screening instruments to identify likely community cases of anxiety and depression (Slade, Grove and Burgess, 2011[43]). Furthermore, they have been extensively validated, including in cross-cultural settings, against diagnostic interview evaluations of anxiety and affective disorders, with lesser but significant associations with other mental disorder categories and with the presence of any current mental disorder (Andrews and Slade, 2001[44]). There is also some evidence that the Kessler scales can be used successfully (with lower cut-off scoring criteria) to capture individuals struggling with more moderate psychological distress that nonetheless warrants mental health intervention (Prochaska et al., 2012[45]).

Table 2.13. Kessler Scale 10/6 Questionnaire with scoring breakdown

	None of the time	A little of the time	Some of the time	Most of the time	All of the time
During the last 30 days, about how often did you feel:					
1. Tired out for no good reason?	1	2	3	4	5
2. Nervous?	1	2	3	4	5
3. So nervous that nothing could calm you down?	1	2	3	4	5
4. Hopeless?	1	2	3	4	5
5. Restless or fidgety?	1	2	3	4	5
6. So restless you could not sit still?	1	2	3	4	5
7. Depressed?	1	2	3	4	5
8. That everything was an effort?	1	2	3	4	5
9. So sad that nothing could cheer you up?	1	2	3	4	5
10. Worthless?	1	2	3	4	5

Note: All items are added together to provide a total score, where higher values indicate worse mental health. However, different scoring methods for both K10 and K6 scales have been used depending on the country and institutional context. For instance, in the United States, answers are coded from 0-4 (leading to a maximum possible score of 40 for the K10 and 24 for the K6), whereas in Australia, 1-5 as shown in the table above have been used (leading to a maximum possible score of 50 for the K10 and 30 for the K6). The K10 scoring used in Australian health surveys have typically been as follows: 10-5 low, 16-21 moderate, 22-29 high, 30-50 very high psychological distress. For the K6 scoring, respondents with scores of 13 (in the 0-4 coding)/ 19 (in the 1-5 coding) or higher are typically classified as having a probable serious mental illness. Cut-off scores in other contexts might vary.
Source: ABS (2007[46]), *Information Paper: Use of the Kessler Psychological Distress Scale in ABS Health Surveys*, Australian Bureau of Statistics, https://www.abs.gov.au/ausstats/abs@.nsf/lookup/4817.0.55.001chapter92007-08; Kessler, R. et al. (2010[47]), "Screening for serious mental illness in the general population with the K6 screening scale: Results from the WHO World Mental Health (WMH) survey initiative", *International Journal of Methods in Psychiatric Research*, Vol. 19/S1, pp. 4-22, https://doi.org/10.1002/mpr.310.

General Health Questionnaire (GHQ-12): The 12-item General Health Questionnaire (GHQ-12) is a measure to detect psychological distress by focusing on affect (negative and positive), somatic symptoms and the functional impairment of respondents. The GHQ-12 has been translated into many languages and extensively validated in general and clinical populations worldwide (particularly against depression and anxiety disorders), including among adolescent samples (Hankins, 2008[48]; Gilbody, 2001[49]; Baksheev et al., 2011[50]). Originally intended as a unidimensional measure, there is some debate about the dimensionality of the GHQ-12, with many factor-analytical studies supporting a range of multidimensional structures (e.g. anxiety and depression, social dysfunction, loss of confidence) (Gao et al., 2004[51]). However, more recent evidence points to these results likely being an expression of method-specific variance caused by item wording, supporting the notion that treating the scale as a unitary construct would minimise bias (Hystad and Johnsen, 2020[52]). The GHQ-12 is subject to copyright restrictions and can thus not be republished in this report.

Patient Health Questionnaire (PHQ-9/ PHQ-8): The full Patient Health Questionnaire (PHQ) contains 59 questions, with modules focusing on mood, anxiety, alcohol, eating and somatoform disorders. The PHQ-9 is a nine-question survey designed to detect the presence and severity of depressive symptoms, and it directly maps onto the DSM-IV and DSM-5 symptom criteria for major depressive disorder. The PHQ-8 questionnaire removes the final question regarding suicidal ideation. While a one-factor structure for both the PHQ-8/9 has been identified, more recent studies support a two-factor model composed of affective and somatic factors (Sunderland et al., 2019[53]). Both instruments have shown acceptable diagnostic screening properties across various population and clinical settings, age groups, and cultures/ ethnicities, in addition to being also a reliable and valid measure of depression severity (Manea, Gilbody and McMillan, 2012[54]; Moriarty et al., 2015[55]; Kroenke et al., 2009[56]; Huang et al., 2006[57]; Kroenke, Spitzer and Williams, 2001[58]; Richardson et al., 2010[59]). The close alignment between the PHQ-8/9 and the DSM make it subject to the same criticism, including a potentially Western-focused construct of

depression, relative to longer self-reported scales with less constrained symptom sets (Zimmerman et al., 2012[60]; Haroz et al., 2017[61]).

Table 2.14. PHQ-9/8 questionnaire with scoring breakdown

	Not at all	Several days	More than half the days	Nearly every day
Over the last two weeks, how often have you been bothered by any of the following problems:				
1. Little interest or pleasure in doing things	0	1	2	3
2. Feeling down, depressed or hopeless	0	1	2	3
3. Trouble falling or staying asleep, or sleeping too much	0	1	2	
4. Feeling tired or having little energy	0	1	2	3
5. Poor appetite or overeating	0	1	2	3
6. Feeling bad about yourself – or that you are a failure or have let yourself or your family down	0	1	2	3
7. Trouble concentrating on things, such as reading the newspaper or watching television	0	1	2	3
8. Moving or speaking so slowly that other people could have noticed. Or the opposite – being so fidgety or restless that you have been moving around a lot more than usual	0	1	2	3
9. Thoughts that you would be better off dead or of hurting yourself in some way	0	1	2	3

Note: The last item in italics is the question on suicidal ideation that is added for the PHQ-9. Scoring can be done in two ways: (1) via an "algorithm diagnosis" of either major depression or other depression; or (2) via summing all items and applying different cut-off scores for depression severity. In the algorithm diagnosis that adheres to DSM definitions, the first or second item (depressed mood or anhedonia) have to present at least "more than half the days" and, combined with at least 5 of the total symptoms or 2 to 4 symptoms also present at this frequency, constitutes major depression or other depression, respectively. In the second form of categorisation, all items are added together to provide a total score of depression severity, with scores ranging from 0-24 for the PHQ-8 and 0-27 for the PHQ-9: 0-4 none, 5-9 mild depression, 10-14 moderate depression, 15-19 moderately severe depression, 20-24/27 severe depression. A score of ≥10 indeed typically represents clinically significant depression regardless of diagnostic status.
Source: Kroenke, K. et al. (2009[56]), "The PHQ-8 as a measure of current depression in the general population", *Journal of Affective Disorders*, Vol. 114/1-3, pp. 163-173, https://doi.org/10.1016/j.jad.2008.06.026; Kroenke, K. et al. (2001[58]), "The PHQ-9: Validity of a brief depression severity measure", *Journal of General Internal Medicine*, Vol. 16/9, pp. 606-613, http://dx.doi.org/10.1046/j.1525-1497.2001.016009606.x.

The Generalised Anxiety Disorder Questionnaire (GAD-7/GAD-2): The Generalised Anxiety Disorder Questionnaire (GAD-7) comprises seven questions about the frequency of broad anxiety-related problems in the past two weeks. It was developed for screening and severity assessment of Generalised Anxiety Disorder, and the items cover most but not all (symptoms of this disorder listed in the DSM-IV and 5 (excessive worry, difficulty to control the worry, restlessness and irritability but not e.g. fatigue, muscle tension, sleep disturbance). Research supports a unidimensional structure for the scale (Sunderland et al., 2019[53]). The GAD-7 has demonstrated good internal consistency, convergent validity, and sensitivity to change in both patient and population samples (Löwe et al., 2008[62]; Beard and Björgvinsson, 2014[63]). While the scale has been successfully translated into multiple languages and local dialects, more research on potential cross-cultural bias of the tool needs to be conducted (Parkerson et al., 2015[64]; Sunderland et al., 2019[53]). The scale focuses on general symptoms of anxiety and was not developed to assess the presence of other anxiety disorders, such as Social Anxiety Disorder. However, some researchers have argued that it can be used across different anxiety disorders, given the scale's emphasis on the transdiagnostic process of worry and the fact that Generalised Anxiety Disorder has a high degree of comorbidity (Johnson et al., 2019[65]). The GAD-2 shorter version of this scale focuses only on the first two items (worry and difficulty to control the worry), i.e. the core criteria of generalised anxiety per the DSM. Available evidence has indicated support for its psychometric properties and validity in a range of settings

(Byrd-Bredbenner, Eck and Quick, 2021[66]; Hughes et al., 2018[67]; Luo et al., 2019[68]; Ahn, Kim and Choi, 2019[69]).

Table 2.15. GAD-7/GAD-2 Questionnaire with scoring breakdown

	Not at all	Several days	More than half the days	Nearly every day
Over the last two weeks, how often have you been bothered by any of the following problems:				
1. Feeling nervous, anxious or on edge	0	1	2	3
2. Not being able to stop or control worrying	0	1	2	3
3. Worrying too much about different things	0	1	2	3
4. Trouble relaxing	0	1	2	3
5. Being so restless that it is hard to sit still	0	1	2	3
6. Becoming easily annoyed or irritable	0	1	2	3
7. Feeling afraid as if something awful might happen	0	1	2	3

Note: Items in italics represent the 2-item shorter version of the scale (GAD-2). All items are added together to provide a total score ranging from 0-21 for the GAD-7, with higher scores indicating the presence of more anxiety symptomatology: 0-4 none, 5-9 mild anxiety, 10-14 moderate anxiety, 15-21 severe anxiety. For the GAD-2, a score of 3 points is the suggested cut-off for identifying possible cases for which further diagnostic evaluation for generalised anxiety disorder is warranted.
Source: Spitzer, R. et al. (2006[70]), "A brief measure for assessing generalized anxiety disorder: The GAD-7", *Archives of Internal Medicine*, Vol. 166/10, pp. 1092-1097, http://dx.doi.org/10.1001/ARCHINTE.166.10.1092.

Patient Health Questionnaire (PHQ-4): The PHQ-4 screening tool is a short, four-question tool to identify the presence and severity of core symptoms of both depression and anxiety, given that these are two of the most prevalent illnesses among the general population and often comorbid. The PHQ-4 pulls the two core depression-related questions from the PHQ-9/8 (which together are called the PHQ-2) plus two core anxiety-related questions from GAD-7 (which are called the GAD-2). Thus, the PHQ-4 is a combination of the PHQ-2 and GAD-2, which have independently been shown to be good, brief screening tools with construct and criterion validity (see above). Available evidence supports the PHQ-4's psychometric properties, reliability and validity in studies focused on the general population, intervention, and workers and college students (Stanhope, 2016[71]; Khubchandani et al., 2016[72]; Löwe et al., 2010[73]).

Table 2.16. PHQ-4 Questionnaire with scoring breakdown

	Not at all	Several days	More than half the days	Nearly every day
Over the last two weeks, how often have you been bothered by any of the following problems:				
1. Feeling nervous, anxious or on edge	0	1	2	3
2. Not being able to stop or control worrying	0	1	2	3
3. Feeling down, depressed or hopeless	0	1	2	3
4. Little interest or pleasure in doing things	0	1	2	3

Note: All items are added together to provide a total score of psychological distress ranging from 0-12, with higher scores indicating the presence of more symptomatology: 0-2 normal, 3-5 mild, 6-8 moderate, 9-12 severe. A total score greater than or equal to 3 for the first two items (GAD-2) indicates that the respondent is at risk for anxiety. A total score greater than or equal to 3 for the final two items (PHQ-2) indicates that the respondent is at risk for depression.
Source: Kroenke, K. et al. (2009[74]), "An ultra-brief screening scale for anxiety and depression: The PHQ-4", *Psychosomatics*, Vol. 50/6, pp. 613-621, http://dx.doi.org/10.1176/APPI.PSY.50.6.613.

Washington Group on Disability Statistics Short Set on Functioning – Enhanced (WG-SS): The Washington Group Short Set on Functioning – Enhanced (WG-SS Enhanced) was developed by the Washington Group on Disability Statistics, which is composed of representatives from National Statistics

Offices, as well as UN agencies, international non-governmental organisations and organisations for people who are disabled, to capture not only the presence but also the type and severity of a respondent's disability for use in population and special interest surveys (Washington Group on Disability Statistics, 2020[75]). Its focus is on functioning in the areas of seeing, hearing, walking or climbing stairs, remembering or concentrating, self-care, communication, upper body activities, as well as affect. The four questions on the latter focus on symptoms of depression and anxiety, though the questionnaire is not typically used in its subcomponent parts. Regardless, the focus on overall functioning might carry important ways forward for capturing transdiagnostic symptoms of mental ill-health.

Table 2.17. WG-SS Enhanced Questionnaire

	No difficulty	Some difficulty	A lot of difficulty	Cannot do at all	
Do you have difficulty:					
1. Seeing, even when wearing your glasses?					
2. Hearing, even when using a hearing aid(s)?					
3. Walking or climbing steps?					
4. Using your usual language, communicating, for example understanding or being understood?					
5. Remembering or concentrating?					
6. With self-care, such as washing all over or dressing?					
7. Raising a 2-liter bottle of water or soda from waist to eye level?					
8. Using your hands and fingers, such as picking up small objects, for example, a button or pencil, or opening or closing containers or bottles?					
	Daily	Weekly	Monthly	A few times a year	Never
9.How often do you feel worried, nervous or anxious?					
	A little	A lot	Somewhere in between a little and a lot		
10. Thinking about the last time you felt worried, nervous or anxious, how would you describe the level of these feelings?					
	Daily	Weekly	Monthly	A few times a year	Never
11. How often do you feel depressed?					
	A little	A lot	Somewhere in between a little and a lot		
12. Thinking about the last time you felt depressed, how depressed did you feel?					

Note: Different domain-specific identifiers of functioning (and the severity of its impairment) can be calculated for an overall disability identifier. The recommended level of inclusion is: "a lot of difficulty" or "cannot do at all" for at least one of the first six questions, severity levels 3 or 4 for the two upper-body mobility questions, and severity level 4 for the anxiety or depression indicators. Items in italics represent the 6-item shorter version of the scale (Washington Group on Disability Statistics Short Set on Functioning), which excludes questions on mental health and upper body functioning.
Source: Washington Group on Disability Statistics (2020[75]), *The Washington Group Short Set on Functioning: Enhanced (WG-SS Enhanced)*, The Washington Group Data Collection Tools and their Recommended Use (washingtongroup-disability.com).

Alcohol Use Disorders Identification Test/Concise (AUDIT/ AUDIT-C): The Alcohol Use Disorders Identification Test (AUDIT) is a 10-item alcohol screen developed by the WHO from the 1980s onwards that can help identify respondents or patients who are hazardous drinkers or have active alcohol use

disorders (including alcohol abuse or dependence). Its validity has been demonstrated in settings beyond primary care, such as inpatient hospital wards, emergency departments, universities, workplaces, outpatient settings and psychiatric services (Berner et al., 2007[76]). Its short version of 3 items, designed to be integrated into routine patient interviews, has been found to have similar accuracy to the full-scale version and has been validated primarily in primary-care settings, as well as increasingly in more general population samples, including adults seeking online help with drinking (Bush et al., 1998[77]; Khadjesari et al., 2017[78]).

Table 2.18. AUDIT/ AUDIT-C Questionnaire with scoring breakdown

	Never	Monthly or less	2-4 times a month	2-3 times a week	4 or more times a week
1. How often do you have a drink containing alcohol?	0	1	2	3	4
	1 or 2	3 to 4	5 to 6	7 to 9	10 or more
2. How many standard drinks containing alcohol do you have on a typical day?	0	1	2	3	4
	Never	Less than monthly	Monthly	Weekly	Daily or almost daily
3. How often do you have six or more drinks on one occasion?	0	1	2	3	4
4. How often during the last year have you found that you were not able to stop drinking once you had started?	0	1	2	3	4
5. How often during the last year have you failed to do what was normally expected from you because of drinking?	0	1	2	3	4
6. How often during the last year have you needed a first drink in the morning to get yourself going after a heavy drinking session?	0	1	2	3	4
7. How often during the last year have you had a feeling of guilt or remorse after drinking?	0	1	2	3	4
8. How often during the last year have you been unable to remember what happened the night before because you had been drinking?	0	1	2	3	4
	No	Yes, but not in the last year	Yes, during the last year		
9. Have you or someone else been injured as a result of your drinking?	0	2	4		
10. Has a relative or friend or a doctor or another health worker been concerned about your drinking or suggested you cut down?	0	2	4		

Note: Items in italics represent the 3-item shorter version of the scale (AUDIT-C). All items are added together to provide a total score ranging from 0-40 (0-12 for the AUDIT-C), with higher scores indicating higher likelihood that a person's drinking is affecting his or her safety. For the AUDIT, scores of 8 or more are recommended as indicators of hazardous and harmful alcohol use, as well as possible alcohol dependence. Since the effects of alcohol vary with average body weight and differences in metabolism, establishing the cut-off point for all women and men over age 65 one point lower at a score of 7 will increase sensitivity for these population groups. For the AUDIT-C, in men (women), a score of 4 (3) or more is considered as identifying symptoms of hazardous drinking or active alcohol use disorders.
Source: Bush, K. et al. (1998[77]), "The AUDIT alcohol consumption questions (AUDIT-C): An effective brief screening test for problem drinking", *Archives of Internal Medicine*, Vol. 158/16, https://doi.org/10.1001/archinte.158.16.1789; WHO (2001[79]), *AUDIT: The Alcohol Use Disorders Identification Test: Guidelines for use in primary health care*, World Health Organization, https://www.who.int/publications/i/item/audit-the-alcohol-use-disorders-identification-test-guidelines-for-use-in-primary-health-care.

Positive mental health

Core questions from the OECD Guidelines on Measuring Subjective Well-being: The *OECD Guidelines on Subjective Well-being* propose a minimal set of measures of subjective well-being covering both life evaluation and (short-term) affect that could be included in household surveys (OECD, 2013[80]). The core measures included are the ones which have the strongest evidence when it comes to validity and relevance, and for which international comparability is the most important. An experimental measure of an aspect of eudaimonic well-being is also included.

Table 2.19. OECD core questions on subjective well-being

	0-10
The following question asks how satisfied you feel, on a scale from 0 to 10. Zero means you feel "not at all satisfied" and 10 means you feel "completely satisfied".	
1. Overall, how satisfied are you with life as a whole these days?	
The following question asks how worthwhile you feel the things you do in your life are, on a scale from 0 to 10. Zero means you feel the things you do in your life are "not at all worthwhile", and 10 means "completely worthwhile".	
2. Overall, to what extent do you feel the things you do in your life are worthwhile?	
The following questions ask about how you felt yesterday on a scale from 0 to 10. Zero means you did not experience the feeling "at all" yesterday while 10 means you experienced the feeling "all of the time" yesterday. I will now read out a list of ways you might have felt yesterday.	
3. How about happy?	
4. How about worried?	
5. How about depressed?	

Note: The three questions on affect (3-5) should be included as a group and are intended to provide a minimal set of questions required to characterise the affective state of the respondent on the previous day.
Source: OECD (2013[80]), *OECD Guidelines on Measuring Subjective Well-being*, OECD Publishing, Paris, https://doi.org/10.1787/9789264191655-en.

WHO-5 Well-being index (WHO-5): The World Health Organization Well-Being Index (WHO-5) is a short questionnaire of 5 items that focus on a respondent's positive affect. The questionnaire, adapted from the longer WHO/ICD-10 Depression Diagnosis and DSM-IV Depression scale by selecting a subset of positively phrased items, has first been used in a project on well-being measures in primary health care by the WHO Regional Office in Europe in 1998 and since then has been translated into more than 30 languages (World Health Organization, 1998[81]; Topp et al., 2015[7]). The WHO-5 has been applied as a generic scale for well-being across a wide range of study fields and countries, as a sensitive screening tool for depression as well as an outcome measure in clinical trials (Topp et al., 2015[7]). Studies of younger and elderly persons indicated a unidimensional structure for this scale (Topp et al., 2015[7]).

Table 2.20. WHO-5 questionnaire with scoring breakdown

	All of the time	Most of the time	More than half the time	Less than half the time	Some of the time	At no time
Over the past two weeks…						
1. I have felt cheerful and in good spirits	5	4	3	2	1	0
2. I have felt calm and relaxed	5	4	3	2	1	0
3. I have felt active and vigorous	5	4	3	2	1	0
4. I woke up feeling fresh and rested	5	4	3	2	1	0
5. My daily life has been filled with things that interest me	5	4	3	2	1	0

Note: All items are added together to provide a total score from 0 to 25, which is then multiplied by 4 to normalise to a 0 (worst possible well-being) to 100 (best possible well-being) score. A cut-off score of less than or equal to 50, or less than or equal to 52 (Sándor et al., 2021[82]),is often used as indicative of reduced well-being, which has been validated in studies using the WHO-5 for the screening of depression and for predicting patient mortality.
Source: Topp, C. et al. (2015[7]), "The WHO-5 well-being index: A systematic review of the literature", *Psychotherapy and Psychosomatics*, Vol. 84/3, pp. 167-176, https://doi.org/10.1159/000376585.

SF-36 Energy/Vitality subscale: The 4-item vitality subscale of the larger SF-36 measure (see above) is a general measure of energy/fatigue. It has been validated in clinical settings and performed well compared to longer scales (e.g. for cancer-related fatigue) (Brown et al., 2011[83]).

Table 2.21. SF-36 vitality subscale questionnaire with scoring breakdown

	All of the time	Most of the time	A good bit of the time	Some of the time	A little of the time	None of the time
How much of the time during the past 4 weeks…						
1. Did you feel full of pep?	1	2	3	4	5	6
2. Did you have a lot of energy?	1	2	3	4	5	6
3. Did you feel worn out?	1	2	3	4	5	6
4. Did you feel tired?	1	2	3	4	5	6

Note: Standardised scores range from 0-100, with lower scores indicating greater fatigue. Scores ≤45 have been established as representing clinically significant fatigue.
Source: Ware, J. et al. (1993[84]), *SF-36 Health Survey: Manual and Interpretation Guide*, The Health Institute, New England Medical Center Hospitals, https://www.researchgate.net/profile/John-Ware-6/publication/313050850_SF-36_Health_Survey_Manual_Interpretation_Guide/links/594a5b83aca2723195de5c3d/SF-36-Health-Survey-Manual-Interpretation-Guide.pdf (accessed on 22 January 2023); Donovan, K. et al. (2008[85]), "Identifying clinically meaningful fatigue with the Fatigue Symptom Inventory", *Journal of Pain and Symptom Management*, Vol. 36/5, pp. 480-487, https://doi.org/10.1016/j.jpainsymman.2007.11.013.

Satisfaction with Life Scale (SWLS): The Satisfaction with Life Scale was developed to assess people's satisfaction and evaluation of their lives as a whole, rather than focusing on specific life domains. Early studies have found it to show good convergent validity with other types of subjective well-being, while being distinct from affective well-being measures (Pavot et al., 1991[86]; Pavot and Diener, 1993[87]).

Table 2.22. SWLS questionnaire with scoring breakdown

	Strongly agree	Agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Disagree	Strongly disagree
1. In most ways my life is close to my ideal.	7	6	5	4	3	2	1
2. The conditions of my life are excellent.	7	6	5	4	3	2	1
3. I am satisfied with my life.	7	6	5	4	3	2	1
4. So far I have gotten the most important things I want in life.	7	6	5	4	3	2	1
5. If I could live my life over, I would change almost nothing.	7	6	5	4	3	2	1

Note: All items are added together to provide a total score from 5 to 35, where higher values indicate higher life satisfaction: 5-9 extremely dissatisfied, 10-14 dissatisfied, 15-19 slightly dissatisfied, 20-24 slightly satisfied, 25-29 satisfied, 30-35 extremely satisfied.
Source: Diener, E. et al. (1985[88]), "The Satisfaction with Life Scale", *Journal of Personality Assessment*, Vol. 49/1, pp. 71-75, https://doi.org/10.1207/s15327752jpa4901_13.

The Mental Health Continuum Short-Form (MHC-SF): The MHC-SF is a 14-item scale developed by Keyes to capture positive mental health in his dual-continuum model (Keyes, 2002[89]). It was derived from

the 40-item Mental Health Continuum Long Form (MHC-LF), and consists of separate subscales: three "emotional well-being" items (reflecting affective well-being plus life satisfaction), five "social well-being" items, and six "psychological well-being" items (which when combined reflect eudaimonic well-being) (Lamers et al., 2011[90]). Studies have shown high internal and moderate test-retest reliability for the MHC-SF and confirmed the 3-factor structure of the subscales, which also show convergent validity with corresponding aspects of well-being and functioning (Lamers et al., 2011[90]).

Table 2.23. MHC-SF questionnaire with scoring breakdown

	Never	Once or twice	About once a week	Two or three times a week	Almost every day	Every day
How often in the past month did you feel …						
Emotional well-being (affect)						
1. Happy?	0	1	2	3	4	5
2. Interested in life?	0	1	2	3	4	5
3. Satisfied with your life?	0	1	2	3	4	5
Social well-being (eudaimonic)						
4. That you had something important to contribute to society? (social contribution)	0	1	2	3	4	5
5. That you belonged to a community (like a social group, your neighbourhood, your city, your school)? (social integration)	0	1	2	3	4	5
6. That our society is becoming a better place for people like you? (social growth)	0	1	2	3	4	5
7. That people are basically good? (social acceptance)	0	1	2	3	4	5
8. That the way our society works makes sense to you? (social coherence)	0	1	2	3	4	5
Psychological well-being (eudaimonic)						
9. That you liked most parts of your personality? (self-acceptance)	0	1	2	3	4	5
10. Good at managing the responsibilities of your daily life? (environmental mastery)	0	1	2	3	4	5
11. That you had warm and trusting relationships with others? (positive relationship with others)	0	1	2	3	4	5
12. That you had experiences that challenged you to grow and become a better person? (personal growth)	0	1	2	3	4	5
13. Confident to think or express your own ideas and opinions? (autonomy)	0	1	2	3	4	5
14. That your life has a sense of direction or meaning to it? (purpose in life)	0	1	2	3	4	5

Note: All items are summed, yielding a total score ranging from 0 to 70, with higher scores indicating greater levels of positive mental health. Subscale scores range from 0 to 15 for emotional well-being, from 0 to 25 for social well-being and from 0 to 30 for psychological well-being. "Flourishing" is defined by reporting ≥ 1 of 3 emotional signs and ≥ 6 of 11 eudaimonic signs (social and psychological subscales combined) experienced "every day" or "almost every day". "Languishing" is defined by reporting ≥ 1 of 3 emotional signs and ≥ 6 of 11 eudaimonic signs experienced "never" or "once or twice". Individuals who are neither flourishing nor languishing are categorised as "moderately mentally healthy".
Source: Lamers, S. et al., (2011[90]), "Evaluating the psychometric properties of the Mental Health Continuum-Short Form (MHC-SF)", *Journal of Clinical Psychology*, Vol. 67/1, pp. 99-110, https://doi.org/10.1002/jclp.2074.

The Warwick-Edinburgh Mental Well-Being Scale* (*WEMWBS): The 14-item WEMWBS scale was developed with funding from NHS Health Scotland in 2005 to measure mental well-being (conceived of as "both feeling good and functioning well"), taking the Affectometer 2 instrument as the starting point (Warwick Medical School, 2021[91]). Some studies confirmed a unidimensional structure for WEMWBS,

while others identified three residual factors relating to affective well-being, psychological functioning or eudaimonia, and social relationships (Shannon et al., 2020[92]; Koushede et al., 2019[93]). A shorter, 7-item version of the scale, SWEMWBS, is also available, focusing slightly less on affect (Stewart-Brown et al., 2009[94]). (S)WEMWBS has been validated in various populations and among different subgroups, including adolescents, clinical samples and ethnic minority samples, and has been translated into more than 25 languages and validated in Norwegian, Swedish, Italian, Dutch, Danish, German, French and Spanish. Both scales have been shown to be sensitive to changes that occur in mental well-being promotion and mental illness treatment and prevention projects (Koushede et al., 2019[93]). Both instruments can distinguish mental well-being between subgroups, but SWEMBS has been found to be less sensitive than the longer version to gender differences (Koushede et al., 2019[93]; Ng Fat et al., 2017[95]).

Table 2.24. (S)WEMWBS questionnaire with scoring breakdown

	None of the time	Rarely	Some of the time	Often	All of the time
Over the last two weeks…					
1. I've been feeling optimistic about the future	1	2	3	4	5
2. I've been feeling useful	1	2	3	4	5
3. I've been feeling relaxed	1	2	3	4	5
4. I've been feeling interested in other people	1	2	3	4	5
5. I've had energy to spare	1	2	3	4	5
6. I've been dealing with problems well	1	2	3	4	5
7. I've been thinking clearly	1	2	3	4	5
8. I've been feeling good about myself	1	2	3	4	5
9 I've been feeling close to other people	1	2	3	4	5
10. I've been feeling confident	1	2	3	4	5
11 I've been able to make up my own mind about things	1	2	3	4	5
12 I've been feeling loved	1	2	3	4	5
13 I've been interested in new things	1	2	3	4	5
14 I've been feeling cheerful	1	2	3	4	5

Note: Items in italics represent the 7-item shorter version of the scale (SWEMWBS). For the 14-item scale, all items are summed, yielding a total score ranging from 14-70. For the 7-item scale, raw scores are transformed into a 7-35 metric score (see conversion table here: https://warwick.ac.uk/fac/sci/med/research/platform/wemwbs/using/howto/swemwbs_raw_score_to_metric_score_conversion_table.pdf). For both scales, higher scores indicate greater levels of positive mental health. (S)WEMWBS scores approximate to a normal distribution, permitting parametric analysis. For categorical scoring, cut-off points for high, average and low mental well-being can be generated using two approaches: (1) a statistical approach putting the cut-off point at +/- one standard deviation, placing approximately 15% of the sample into high well-being and 15% into low well-being categories; or (2) a benchmarking approach against validated measures of depression, e.g. a score of 41-44 as indicative of possible/mild depression and a score of >41 as indicative of probable clinical depression, using the Center for Epidemiologic Studies Depression Scale (CES-D) as a benchmark. WEMWBS is protected by copyright. Those wishing to use WEMWBS can obtain a licence to do so. Please go to https://warwick.ac.uk/wemwbs/using for information on the type of licence you will require and details on how to apply. A free-of-charge "non-commercial" licence is available to public sector organisations, charities and registered social enterprises, as well as to researchers employed in Higher Education Institutions. Any further enquiries can be directed to wemwbs@warwick.ac.uk.

Source: Warwick-Edinburgh Mental Well-being Scale (WEMWBS) © NHS Health Scotland, University of Warwick and University of Edinburgh, 2006, all rights reserved; Warwick Medical School (2021[91]), *The Warwick-Edinburgh Mental Wellbeing Scales (WEMWBS),* https://warwick.ac.uk/fac/sci/med/research/platform/wemwbs/.

References

ABS (2007), *Information Paper: Use of the Kessler Psychological Distress Scale in ABS Health Surveys*, Australian Bureau of Statistics, https://www.abs.gov.au/ausstats/abs@.nsf/lookup/4817.0.55.001chapter92007-08. [46]

Ahn, J., Y. Kim and K. Choi (2019), "The psychometric properties and clinical utility of the Korean version of GAD-7 and GAD-2", *Frontiers in Psychiatry*, Vol. 127/10, https://doi.org/10.3389/fpsyt.2019.00127. [69]

Andrews, G. and T. Slade (2001), "Interpreting scores on the Kessler Psychological Distress Scale (K10)", *Australian and New Zealand Journal of Public Health*, Vol. 25/6, pp. 494-497, https://doi.org/10.1111/j.1467-842X.2001.tb00310.x. [44]

Baksheev, G. et al. (2011), "Validity of the 12-item General Health Questionnaire (GHQ-12) in detecting depressive and anxiety disorders among high school students", *Psychiatry Research*, Vol. 187/1-2, pp. 291-296, https://doi.org/10.1016/j.psychres.2010.10.010. [50]

Beard, C. and T. Björgvinsson (2014), "Beyond generalized anxiety disorder: Psychometric properties of the GAD-7 in a heterogeneous psychiatric sample", *Journal of Anxiety Disorders*, Vol. 28/6, pp. 547-552, https://doi.org/10.1016/j.janxdis.2014.06.002. [63]

Berner, M. et al. (2007), "The Alcohol Use Disorders Identification Test for Detecting At-Risk Drinking: A Systematic Review and Meta-Analysis", *Journal of Studies on Alcohol and Drugs*, Vol. 68/3, pp. 461-473, https://doi.org/10.15288/jsad.2007.68.461. [76]

Berwick, D. et al. (1991), "Performance of a five-item mental health screening test", *Medical Care*, Vol. 29/2, pp. 169-176, https://doi.org/10.1097/00005650-199102000-00008. [28]

Brooks, R., J. Beard and Z. Steel (2006), "Factor structure and interpretation of the K10", *Psychological Assessment*, Vol. 18/1, pp. 62-70, https://doi.org/10.1037/1040-3590.18.1.62. [42]

Brown, L. et al. (2011), "Comparison of SF-36 vitality scale and Fatigue Symptom Inventory in assessing cancer-related fatigue", *Supportive Care in Cancer*, Vol. 19/8, pp. 1255-1259, https://doi.org/10.1007/s00520-011-1148-2. [83]

Burger, H. and J. Neeleman (2007), "A glossary on psychiatric epidemiology", *Journal of Epidemiology and Community Health*, Vol. 61/3, pp. 185-189, https://doi.org/10.1136/jech.2003.019430. [4]

Bush, K. et al. (1998), "The AUDIT alcohol consumption questions (AUDIT-C): An effective brief screening test for problem drinking", *Archives of Internal Medicine*, Vol. 158/16, https://doi.org/10.1001/archinte.158.16.1789. [77]

Byrd-Bredbenner, C., K. Eck and V. Quick (2021), "GAD-7, GAD-2, and GAD-mini: Psychometric properties and norms of university students in the United States", *General Hospital Psychiatry*, Vol. 69, pp. 61-66, https://doi.org/10.1016/j.genhosppsych.2021.01.002. [66]

Case, A. and A. Deaton (2017), *Mortality and Morbidity in the 21st Century*, Brookings Papers on Economic Activity, http://brookings.edu/wp-content/uploads/2017/08/casetextsp17bpea.pdf. [2]

Chile, G. (2021), *Saludable Mente [Healthy Mind]*, https://www.gob.cl/saludablemente/. [21]

Diener, E. et al. (1985), "The Satisfaction with Life Scale", *Journal of Personality Assessment*, Vol. 49/1, pp. 71-75, https://doi.org/10.1207/s15327752jpa4901_13. [88]

Dobson, K. et al. (2020), "Trends in the prevalence of depression and anxiety disorders among Canadian working-age adults between 2000 and 2016", *Health Reports*, Vol. 31/12, pp. 12-23, https://doi.org/10.25318/82-003-X202001200002-ENG. [9]

Donovan, K. et al. (2008), "Identifying clinically meaningful fatigue with the fatigue symptom inventory", *Journal of Pain and Symptom Management*, Vol. 36/5, pp. 480-487, https://doi.org/10.1016/j.jpainsymman.2007.11.013. [85]

Elovanio, M. et al. (2020), "General Health Questionnaire (GHQ-12), Beck Depression Inventory (BDI-6), and Mental Health Index (MHI-5): Psychometric and predictive properties in a Finnish population-based sample", *Psychiatry Research*, Vol. 289, p. 112973, https://doi.org/10.1016/j.psychres.2020.112973. [32]

Gao, F. et al. (2004), "Does the 12-item General Health Questionnaire contain multiple factors and do we need them?", *Health and Quality of Life Outcomes*, Vol. 2/1, p. 63, https://doi.org/10.1186/1477-7525-2-63. [51]

Gilbody, S. (2001), "Routinely administered questionnaires for depression and anxiety: systematic review", *BMJ*, Vol. 322/7283, pp. 406-409, https://doi.org/10.1136/bmj.322.7283.406. [49]

Gill, S. et al. (2007), "Validity of the mental health component scale of the 12-item Short-Form Health Survey (MCS-12) as measure of common mental disorders in the general population", *Psychiatry Research*, Vol. 152/1, pp. 63-71, https://doi.org/10.1016/j.psychres.2006.11.005. [33]

Hankins, M. (2008), "The reliability of the twelve-item general health questionnaire (GHQ-12) under realistic assumptions", *BMC Public Health*, Vol. 8/1, p. 355, https://doi.org/10.1186/1471-2458-8-355. [48]

Haroz, E. et al. (2017), "How is depression experienced around the world? A systematic review of qualitative literature", *Social Science & Medicine*, Vol. 183, pp. 151-162, https://doi.org/10.1016/j.socscimed.2016.12.030. [61]

Harvard Medical School (n.d.), *About the WHO WMH-CIDI*, https://www.hcp.med.harvard.edu/wmhcidi/about-the-who-wmh-cidi/ (accessed on 31 May 2022). [27]

HHS (2021), *U.S. Surgeon General Issues Advisory on Youth Mental Health Crisis Further Exposed by COVID-19 Pandemic*, United States Department of Health and Human Services (HHS), https://www.hhs.gov/about/news/2021/12/07/us-surgeon-general-issues-advisory-on-youth-mental-health-crisis-further-exposed-by-covid-19-pandemic.html. [20]

Hoeymans, N. et al. (2004), "Measuring mental health of the Dutch population: A comparison of the GHQ-12 and the MHI-5", *Health and Quality of Life Outcomes*, Vol. 2/23, https://doi.org/10.1186/1477-7525-2-2. [31]

Huang, F. et al. (2006), "Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients", *Journal of General Internal Medicine*, Vol. 21/6, pp. 547-552, https://doi.org/10.1111/j.1525-1497.2006.00409.x. [57]

Hughes, A. et al. (2018), "Diagnostic and clinical utility of the GAD-2 for screening anxiety symptoms in individuals with multiple sclerosis", *Archives of Physical Medicine and Rehabilitation*, Vol. 99/10, pp. 2045-2049, https://doi.org/10.1016/j.apmr.2018.05.029. [67]

Hystad, S. and B. Johnsen (2020), "The Dimensionality of the 12-Item General Health Questionnaire (GHQ-12): Comparisons of Factor Structures and Invariance Across Samples and Time", *Frontiers in Psychology*, Vol. 11, https://doi.org/10.3389/fpsyg.2020.01300. [52]

Johnson, S. et al. (2019), "Psychometric properties of the General Anxiety Disorder 7-Item (GAD-7) scale in a heterogeneous psychiatric sample", *Frontiers in Psychology*, Vol. 10, https://doi.org/10.3389/fpsyg.2019.01713. [65]

Kelly, M. et al. (2008), "Evaluating cutpoints for the MHI-5 and MCS using the GHQ-12: A comparison of five different methods", *BMC Psychiatry*, Vol. 8/10, https://doi.org/10.1186/1471-244X-8-10. [29]

Kessler, R. and T. Bedirhan Üstün (2006), "The World Mental Health (WMH) survey initiative version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI)", *International Journal of Methods in Psychiatric Research*, Vol. 13/2, pp. 93-121, https://doi.org/10.1002/mpr.168. [26]

Kessler, R. et al. (2010), "Screening for serious mental illness in the general population with the K6 screening scale: results from the WHO World Mental Health (WMH) survey initiative", *International Journal of Methods in Psychiatric Research*, Vol. 19/S1, pp. 4-22, https://doi.org/10.1002/mpr.310. [47]

Keyes, C. (2002), "The mental health continuum: From languishing to flourishing in life", *Journal of Health and Social Behavior*, Vol. 43/2, pp. 207-222, https://doi.org/10.2307/3090197. [89]

Khadjesari, Z. et al. (2017), "Validation of the AUDIT-C in adults seeking help with their drinking online", *Addiction Science & Clinical Practice*, Vol. 12/1, p. 2, https://doi.org/10.1186/s13722-016-0066-5. [78]

Khubchandani, J. et al. (2016), "The psychometric properties of PHQ-4 depression and anxiety screening scale among college students", *Archives of Psychiatric Nursing*, Vol. 30/4, pp. 457-462, https://doi.org/10.1016/j.apnu.2016.01.014. [72]

KOSIS (n.d.), *Annual prevalence of mental disorders (adjusted for sex and age) (database)*, Korean Statistical Information Service, https://kosis.kr/statHtml/statHtml.do?orgId=117&tblId=TX_117_2009_HB027&conn_path=I2 (accessed on 16 May 2022). [10]

KOSIS (n.d.), *Depressive disorder prevalence (database)*, National Health and Nutrition Survey, Korean Statistical Information Service, https://knhanes.kdca.go.kr/knhanes/sub01/sub01_05.do#none (accessed on 31 August 2022). [11]

Koushede, V. et al. (2019), "Measuring mental well-being in Denmark: Validation of the original and short version of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS and SWEMWBS) and cross-cultural comparison across four European settings", *Psychiatry Research*, Vol. 271, pp. 502-509, https://doi.org/10.1016/j.psychres.2018.12.003. [93]

Kroenke, K., R. Spitzer and J. Williams (2001), "The PHQ-9: Validity of a brief depression severity measure", *Journal of General Internal Medicine*, Vol. 16/9, https://doi.org/10.1046/j.1525-1497.2001.016009606.x. [58]

Kroenke, K. et al. (2009), "An ultra-brief screening scale for anxiety and depression: The PHQ-4", *Psychosomatics*, Vol. 50/6, pp. 613-621, https://doi.org/10.1176/APPI.PSY.50.6.613. [74]

Kroenke, K. et al. (2009), "The PHQ-8 as a measure of current depression in the general population", *Journal of Affective Disorders*, Vol. 114/1-3, pp. 163-173, https://doi.org/10.1016/j.jad.2008.06.026. [56]

Lamers, S. et al. (2011), "Evaluating the psychometric properties of the Mental Health Continuum-Short Form (MHC-SF)", *Journal of Clinical Psychology*, Vol. 67/1, pp. 99-110, https://doi.org/10.1002/jclp.20741. [90]

Löwe, B. et al. (2008), "Validation and standardization of the Generalized Anxiety Disorder screener (GAD-7) in the general population", *Medical Care*, Vol. 46/3, pp. 266-274, https://doi.org/10.1097/mlr.0b013e318160d093. [62]

Löwe, B. et al. (2010), "A 4-item measure of depression and anxiety: Validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population", *Journal of Affective Disorders*, Vol. 122/1-2, pp. 86-95, https://doi.org/10.1016/j.jad.2009.06.019. [73]

Luo, Z. et al. (2019), "Adaptation of the two-item generalized anxiety disorder scale (GAD-2) to Chinese rural population: A validation study and meta-analysis", *General Hospital Psychiatry*, Vol. 60, pp. 50-56, https://doi.org/10.1016/j.genhosppsych.2019.07.008. [68]

Manea, L., S. Gilbody and D. McMillan (2012), "Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): A meta-analysis", *Canadian Medical Association Journal*, Vol. 184/3, pp. E191-E196, https://doi.org/10.1503/cmaj.110829. [54]

Moriarty, A. et al. (2015), "Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): A meta-analysis", *General Hospital Psychiatry*, Vol. 37/6, pp. 567-576, https://doi.org/10.1016/j.genhosppsych.2015.06.012. [55]

Mueller, A. and D. Segal (2015), "Structured versus Semistructed versus Unstructured Interviews", in Cautin, R. and S. Lilienfeld (eds.), *The Encyclopedia of Clinical Psychology*, John Wiley & Sons, Inc, https://www.researchgate.net/profile/Daniel-Segal-6/publication/313966231_Structured_versus_Semistructured_versus_Unstructured_Interviews/links/599a13bcaca272e41d3ec727/Structured-versus-Semistructured-versus-Unstructured-Interviews.pdf (accessed on 30 March 2022). [3]

National Academies of Sciences Engineering and Medicine (2021), *Mental Health, Substance Use, and Wellbeing in Higher Education: Supporting the Whole Student*, The National Academies Press, Washington, D.C., https://doi.org/10.17226/26015. [6]

National Center for Health Statistics (2021), *Estimates of Mental Health Symptomatology, by Month of Interview: United States, 2019 (database)*, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, https://www.cdc.gov/nchs/data/nhis/mental-health-monthly-508.pdf (accessed on 16 May 2022). [13]

Ng Fat, L. et al. (2017), "Evaluating and establishing national norms for mental wellbeing using the short Warwick–Edinburgh Mental Well-being Scale (SWEMWBS): Findings from the Health Survey for England", *Quality of Life Research*, Vol. 26/5, pp. 1129-1144, https://doi.org/10.1007/s11136-016-1454-8. [95]

OECD (2021), *A New Benchmark for Mental Health Systems: Tackling the Social and Economic Costs of Mental Ill-Health*, OECD Health Policy Studies, OECD Publishing, Paris, https://doi.org/10.1787/4ed890f6-en. [1]

OECD (2021), *COVID-19 and Well-being: Life in the Pandemic*, OECD Publishing, Paris, https://doi.org/10.1787/1e1ecb53-en. [18]

OECD (2021), *Health at a Glance*, OECD Publishing, Paris, https://doi.org/10.1787/19991312. [17]

OECD (2021), *Supporting young people's mental health through the COVID-19 crisis*, OECD Policy Responses to Coronavirus (COVID-19), OECD Publishing, Paris, https://www.oecd.org/coronavirus/policy-responses/supporting-young-people-s-mental-health-through-the-covid-19-crisis-84e143e5/. [19]

OECD (2020), *How's Life? 2020: Measuring Well-being*, OECD Publishing, Paris, https://doi.org/10.1787/9870c393-en. [16]

OECD (2013), *OECD Guidelines on Measuring Subjective Well-being*, OECD Publishing, Paris, https://doi.org/10.1787/9789264191655-en. [80]

OECD/European Union (2018), *Health at a Glance: Europe 2018: State of Health in the EU Cycle*, OECD Publishing, Paris/European Union, Brussels, https://doi.org/10.1787/health_glance_eur-2018-en. [23]

Ohrnberger, J. et al. (2020), "Validation of the SF12 mental and physical health measure for the population from a low-income country in sub-Saharan Africa", *Health and Quality of Life Outcomes*, Vol. 18/78, https://doi.org/10.1186/s12955-020-01323-1. [40]

Parkerson, H. et al. (2015), "Cultural-based biases of the GAD-7", *Journal of Anxiety Disorders*, Vol. 31, pp. 38-42, https://doi.org/10.1016/j.janxdis.2015.01.005. [64]

Pavot, W. and E. Diener (1993), "Review of the Satisfaction With Life Scale.", *Psychological Assessment*, Vol. 5/2, pp. 164-172, https://doi.org/10.1037/1040-3590.5.2.164. [87]

Pavot, W. et al. (1991), "Further validation of the Satisfaction with Life Scale: evidence for the cross-method convergence of well-being measures", *Journal of Personality Assessment*, Vol. 57/1, pp. 149-161, https://doi.org/10.1207/s15327752jpa5701_17. [86]

Prochaska, J. et al. (2012), "Validity study of the K6 scale as a measure of moderate mental distress based on mental health treatment need and utilization", *International Journal of Methods in Psychiatric Research*, Vol. 21/2, pp. 88-97, https://doi.org/10.1002/mpr.1349. [45]

Public Health Agency Sweden (2022), *Synen på psykisk ohälsa och suicid*, https://www.folkhalsomyndigheten.se/publicerat-material/publikationsarkiv/s/synen-pa-psykisk-ohalsa-och-suicid-/?pub=105538#105580. [25]

Quality Metric (n.d.), *The SF-12v2 PRO Health Survey*, https://www.qualitymetric.com/health-surveys-old/the-sf-12v2-health-survey/ (accessed on 10 May 2022). [41]

RAND (n.d.), *Mental Health Inventory Survey*, https://www.rand.org/health-care/surveys_tools/mos/mental-health.html (accessed on 10 May 2022). [38]

Richardson, L. et al. (2010), "Evaluation of the Patient Health Questionnaire-9 item for detecting major depression among adolescents", *Pediatrics*, Vol. 126/6, pp. 1117-1123, https://doi.org/10.1542/peds.2010-0852. [59]

Ruedgers, R. (2001), *Handbook of diagnositc and structured interviewing*, Guilford Press, https://psycnet.apa.org/record/2001-05047-000. [5]

Rumpf, H. et al. (2001), "Screening for mental health: validity of the MHI-5 using DSM-IV Axis I psychiatric disorders as gold standard", *Psychiatry Research*, Vol. 105/3, pp. 243-253, https://doi.org/10.1016/S0165-1781(01)00329-8. [34]

SAMHSA (2019), *Key Substance Use and Mental Health Indicators in the United States: Results from the 2018 National Survey on Drug Use and Health (database)*, Substance Abuse and Mental Health Services Administration, Rockville, MD, https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHNationalFindingsReport2018/NSDUHNationalFindingsReport2018.pdf (accessed on 16 May 2022). [12]

Sándor, E. et al. (2021), *Impact of COVID-19 on young people in the EU*, Publications Office of the European Union, https://www.eurofound.europa.eu/publications/report/2021/impact-of-covid-19-on-young-people-in-the-eu#:~:text=COVID%2D19%20pandemic.-,Young%20people%20were%20more%20likely%20than%20older%20groups%20to%20experience,home%20requirements%20and%20school%20closures. [82]

Santé Publique France (2021), *La santé mentale au temps de la COVID-19 : En parler, c'est déjà se soigner [Mental health in the time of COVID-19: Talking about it is already treating yourself]*, Santé Puique France, https://www.santepubliquefrance.fr/presse/2021/la-sante-mentale-au-temps-de-la-covid-19-en-parler-c-est-deja-se-soigner. [22]

Santomauro, D. et al. (2021), "Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic", *The Lancet*, Vol. 398/10312, pp. 1700-1712, https://doi.org/10.1016/S0140-6736(21)02143-7. [24]

Shannon, S. et al. (2020), "Testing the factor structure of the Warwick-Edinburgh Mental Well-Being Scale in adolescents: A bi-factor modelling methodology", *Psychiatry Research*, Vol. 293, https://doi.org/10.1016/j.psychres.2020.113393. [92]

Slade, T., R. Grove and P. Burgess (2011), "Kessler Psychological Distress Scale: Normative Data from the 2007 Australian National Survey of Mental Health and Wellbeing", *Australian & New Zealand Journal of Psychiatry*, Vol. 45/4, pp. 308-316, https://doi.org/10.3109/00048674.2010.543653. [43]

Spitzer, R. et al. (2006), "A brief Measure for assessing Generalized Anxiety Disorder: The GAD-7", *Archives of Internal Medicine*, Vol. 166/10, pp. 1092-1097, https://doi.org/10.1001/ARCHINTE.166.10.1092. [70]

Stanhope, J. (2016), "Patient Health Questionnaire-4", *Occupational Medicine*, Vol. 66/9, pp. 760-761, https://doi.org/10.1093/occmed/kqw165. [71]

Statistics Canada (2013), *Canadian Community Health Survey: Mental Health, 2012*, The Daily, https://www150.statcan.gc.ca/n1/daily-quotidien/130918/dq130918a-eng.htm (accessed on 16 May 2022). [8]

Stewart-Brown, S. et al. (2009), "Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a Rasch analysis using data from the Scottish Health Education Population Survey", *Health and Quality of Life Outcomes*, Vol. 7/1, p. 15, https://doi.org/10.1186/1477-7525-7-15. [94]

Strand, B. et al. (2003), "Measuring the mental health status of the Norwegian population: A comparison of the instruments SCL-25, SCL-10, SCL-5 and MHI-5 (SF-36)", *Nordic Journal of Psychiatry*, Vol. 57/2, pp. 113-118, https://doi.org/10.1080/08039480310000932. [35]

Sunderland, M. et al. (2019), "Self-Report Scales for Common Mental Disorders", in *The Cambridge Handbook of Clinical Assessment and Diagnosis*, Cambridge University Press, https://doi.org/10.1017/9781108235433.019. [53]

Thorsen, S. et al. (2013), "The predictive value of mental health for long-term sickness absence: the Major Depression Inventory (MDI) and the Mental Health Inventory (MHI-5) compared", *BMC Medical Research Methodology*, Vol. 13/115, https://doi.org/10.1186/1471-2288-13-115. [36]

Topp, C. et al. (2015), "The WHO-5 well-being index: A systematic review of the literature", *Psychotherapy and Psychosomatics*, Vol. 84/3, pp. 167-176, https://doi.org/10.1159/000376585. [7]

University of Essex, Institute for Social and Economic Research (2022), *Understanding Society: Waves 1-11, 2009-2020 and Harmonised BHPS: Waves 1-18, 1991-2009 (database)*, 5th Edition. UK Data Service., https://doi.org/10.5255/UKDA-SN-6614-16 (accessed on 10 June 2022). [15]

Vilagut, G. et al. (2013), "The mental component of the Short-Form 12 health survey (SF-12) as a measure of depressive disorders in the general population: Results with three alternative scoring methods", *Value in Health*, Vol. 16/4, pp. 564-573, https://doi.org/10.1016/j.jval.2013.01.006. [39]

Ware, J. et al. (1993), *SF-36 Health Survey: Manual and Interpretation Guide*, The Health Institute, New England Medical Center Hospitals, https://www.researchgate.net/profile/John-Ware-6/publication/313050850_SF-36_Health_Survey_Manual_Interpretation_Guide/links/594a5b83aca2723195de5c3d/SF-36-Health-Survey-Manual-Interpretation-Guide.pdf (accessed on 22 January 2023). [84]

Ware, J. et al. (2002), *How to score SF-12 items: How to score version 2 of the SF-12 Health Survey*, https://www.researchgate.net/publication/291994160. [37]

Warwick Medical School (2021), *The Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS)*, https://warwick.ac.uk/fac/sci/med/research/platform/wemwbs/. [91]

Washington Group on Disability Statistics (2020), *The Washington Group Short Set on Functioning: Enhanced (WG-SS Enhanced)*, https://www.washingtongroup-disability.com/fileadmin/uploads/wg/Documents/Washington_Group_Questionnaire__3_-_WG_Short_Set_on_Functioning_-_Enhanced.pdf. [75]

Woodhead, C. et al. (2012), "Impact of exposure to combat during deployment to Iraq and Afghanistan on mental health by gender", *Psychological Medicine*, Vol. 42/9, pp. 1985-1996, https://doi.org/10.1017/S003329171100290X. [14]

World Health Organization (2001), *AUDIT: The Alcohol Use Disorders Identification Test: Guidelines for use in primary health care*, World Health Organization, https://www.who.int/publications/i/item/audit-the-alcohol-use-disorders-identification-test-guidelines-for-use-in-primary-health-care. [79]

World Health Organization (1998), *Wellbeing measures in primary health care/the DepCare Project: Report on a WHO meeting*, World Health Organization, https://apps.who.int/iris/handle/10665/349766. [81]

Yamazaki, S., S. Fukuhara and J. Green (2005), "Usefulness of five-item and three-item Mental Health Inventories to screen for depressive symptoms in the general population of Japan", *Health and Quality of Life Outcomes*, Vol. 3/1, pp. 1-7, https://doi.org/10.1186/1477-7525-3-48/FIGURES/1. [30]

Zimmerman, M. et al. (2012), "How can we use depression severity to guide treatment selection when measures of depression categorize patients differently?", *Journal of Clinical Psychiatry*, Vol. 73/10, pp. 1287-1291, https://doi.org/10.4088/JCP.12m07775. [60]

Notes

[1] Of course, this implies that diagnoses reached through clinical interviews are only as valid as the classification system they are based on (Mueller and Segal, 2015[3]) (see also Box 3.4 in Chapter 3).

[2] Of course, the coverage of household surveys is also not complete and includes only those sampled. Typically, people living in institutional settings as well as the homeless (who are likely to have higher prevalence of mental ill-health than the general population) are not taken into account.

[3] The following countries responded to the questionnaire: Australia, Austria, Belgium, Bulgaria, Canada, Chile, Colombia, Costa Rica, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, Luxembourg, Mexico, the Netherlands, Norway, New Zealand, Poland, Portugal, Slovenia, the Slovak Republic, Spain, Sweden, Switzerland, Türkiye, the United Kingdom and the United States.

[4] The OECD also publishes administrative data on mental health service provision, such as the number of psychiatrists, psychologists or mental health professionals per 100 000 population; the number of hospital beds devoted to mental health care; spending on mental health services; etc. (OECD, 2021[17]). As these are not considered population-level mental health *outcomes*, they are not further considered for the purposes of this project.

[5] Percentages do not add up to 68% because some countries did both: introduced new stand-alone surveys *and* added mental health modules to existing surveys.

[6] Furthermore, it is worth noting that while the MHI-5 appeared in the well-being ad hoc modules for the 2013 and 2018 European Union Statistics on Income and Living Conditions (EU-SILC) survey administered by Eurostat, in future well-being modules the tool has been removed. Therefore, future use of the MHI-5 may be significantly diminished, although some individual member states may elect to keep the measure in their own national health and/or well-being surveys.

[7] For an extended discussion of surveys used to measure attitudes and stigma towards mental health, refer to Table 6.2 in (OECD, 2021[1]).

3 Good practices for measuring population mental health in household surveys

All OECD countries currently measure population mental health, yet use a variety of tools to capture a multitude of outcomes. In order to improve harmonisation, this chapter poses a series of questions that highlight the criteria to be considered when choosing appropriate survey tools. These criteria include statistical quality, practicalities of fieldwork and data analysis. Overall, there is strong evidence supporting the statistical properties of the most commonly used screening tools for the composite scales of mental ill-health and positive mental health. Four concrete tools (the PHQ-4, the WHO-5 or SWEMWBS, and a question on general mental health status) that capture outcomes across the mental health spectrum are suggested for inclusion in household surveys in addition to already ongoing data collection efforts.

Countries across the OECD are already implementing a variety of survey tools to measure aspects of population mental health. Chapter 2 highlighted that while there is some degree of harmonisation for outcomes such as risk for depression, life evaluation and general psychological distress, there are gaps in coverage for others: in particular, anxiety; other specific mental disorders (bipolar disorder, PTSD, eating disorders and so on); and affect and eudaimonic aspects of positive mental health. Before settling on a concrete list of recommendations for member countries, this chapter provides an overview of the properties that should be considered when selecting a specific tool to measure these outcomes in household surveys.

While OECD countries are already using a variety of tools, including structured interviews, data on previous diagnoses, experienced symptoms and questions on suicidal ideation and suicide attempts, this chapter will focus on the statistical qualities of three other tools in common use: screening tools[1], positive mental health indicators and general questions on mental health status. This is for two reasons. First, these tools are standardised in terms of question formulation and thus provide the easiest foundation on which to make harmonised recommendations. Second, these tools are more commonly featured in general social surveys (as compared to tools for diagnoses or experienced symptoms), which tend to be collected more frequently than health-specific surveys. Taken together, these three tools also provide a holistic measure of mental health, encompassing the full possibility of outcomes conceptualised by both the single and dual continua (see Figure 1.2), and they provide more nuance than, say, measures of suicidal ideation or attempts. These tools are then the most promising when thinking of pragmatic recommendations that can be taken up by the largest number of countries.

When selecting an appropriate tool, the overarching consideration is how to measure the different facets of mental health most accurately – across countries, groups and time – in a way that can be used by government as a part of an integrated policy approach to mental health. High-quality data are needed to provide insights into how societal conditions (economic, social, environmental) affect the mental health of different population groups and whether these conditions contribute to improving or declining mental health. No data are completely without measurement bias, and it is always important that data collection entities enact rigorous quality controls to minimise the amount of noise in a measure. However, there are challenges specific to the measurement of mental health, due to stigma and bias affecting survey response behaviour, different cultural views and evolving attitudes towards mental health over time. Furthermore, household surveys by definition exclude institutionalised populations, including those in long-term care facilities, hospitals or prisons, as well as people with no permanent addresses, all of whom may have higher-than-average risk for some mental health conditions.

Good practices for measuring mental health at the population level differ in several ways from those for measuring mental health at the clinical level. For national statistical offices or health ministries conducting large-scale, nationally representative surveys, implementing long structured interviews is impractical, even though these may be considered the gold standard from a clinical perspective. The end users of the data are different, and policy makers have other needs than clinicians: tracking overall trends (over time, across at-risk groups, among countries), and factors of risk and resilience in population groups vs. diagnosing an individual and developing a treatment plan. These needs guide this chapter's discussion.

This chapter provides a guide to good practices in producing high-quality data on population mental health outcomes, by posing a series of questions for data collectors to consider. High-level findings from this exercise are shown in Table 3.1, below. The specific screening and composite-scale tools included in the table are those that are used most frequently across OECD countries (for more information on each, refer to Table 2.7, Table 2.11, Table 2.12 and Annex 2.B).[2] Questions are grouped into three overarching categories, covering (1) statistical quality, (2) data collection procedures and (3) analysis. Evidence from existing research is used to illustrate each question area, rather than to comprehensively assess every mental health tool used by OECD countries. These framing questions serve as a lens for assessing the advantages and disadvantages of different tools for measuring population mental health and to guide the concrete recommendations for tool take-up and harmonisation outlined in the conclusion.

Table 3.1. Overview of mental health tool performance on statistical quality, data collection and analysis metrics

Tool Information			Statistical Quality					Data Analysis		Country Coverage
Name	Topic coverage and item length	Reference period	Reliability	Validity	Low missing rates	High comparability across groups	Sensitive to change	Normal distribution	Sensitivity/ specificity of thresholds	OECD countries reporting its use
Validated screening tools for assessing mental ill-health										
Psychological distress										
General Health Questionnaire (GHQ-12)	Negative and positive affect, somatic symptoms, functional impairment; 12 items	Recently	✓	×	×	~		×	×	5 of 37
Kessler Scale 6 (K6)	Negative affect; 6 items	Past 4 weeks	◯	✓		~			✓	4 of 37
Kessler Scale 10 (K10)	Negative affect, functional impairment; 10 items	Past 4 weeks	✓	✓		~			✓	4 of 37
Mental Health Inventory 5 (MHI-5)	Negative and positive affect; 5 items	Past month	✓	✓				×	~	28 of 37
Depressive symptoms										
Patient Health Questionnaire -8 or -9 (PHQ-8 / PHQ-9)	Negative affect, anhedonia, somatic symptoms, functional impairment (matched to major depressive disorder per DSM-IV and DSM-5 criteria); 8 or 9 items	Past 2 weeks	✓	✓		✓	✓		✓	30 of 37
Patient Health Questionnaire -2 (PHQ-2)	Negative affect, anhedonia; 2 items	Past 2 weeks	~	~		~	✓			8 of 37
Center for Epidemiological Studies Depression Scale (CES-D)	Negative affect, anhedonia; 20 items	Past week	✓	~		×			×	1 of 37
Symptoms of anxiety										
Generalised Anxiety Disorder-7 (GAD-7)	Negative affect, somatic symptoms, functional impairment; 7 items	Past 2 weeks	✓	✓		~	◯		~	11 of 37

Generalised Anxiety Disorder-2 (GAD-2)	Negative affect, functional impairment; 2 items	Past 2 weeks	✓	✓		~			~	7 of 37
Symptoms of depression and anxiety										
Patient Health Questionnaire -4 (PHQ-4)	Negative affect, anhedonia, functional impairment; 4 items	Past 2 weeks	~	~		~	~		~	13 of 37
Standardised tools for assessing positive mental health										
Short Form Health Status (SF-12)	Negative and positive affect, functional impairment (Mental Health Component Summary); 12 items	Past 4 weeks	✓	✓		×				8* of 37
Warwick- Edinburgh Mental Well-Being Scale (WEMWBS)	Positive affect, eudaimonia, social well-being; 14 items	Past 2 weeks	✓	✓	✓	~	✓	✓	~	2 of 37
Short Warwick-Edinburgh Mental Well-Being Scale (SWEMWBS)	Positive affect, eudaimonia, social well-being; 7 items	Past 2 weeks	✓	✓	✓	✓	✓	✓	~	6 of 37
WHO-5 Wellbeing Index (WHO-5)	Positive affect; 5 items	Past 2 weeks	✓	✓		✓				6 of 37
Mental Health Continuum Short-Form (MHC-SF)	Positive affect, eudaimonia, life satisfaction, social well-being 14 items	Past month	✓	~		~			×	2 of 37
Single-question self-reported general mental health status										
Self-reported mental health (SRMH)	Varies widely, including self-reported: general mental health status; number of mentally healthy days; recovery from mental health condition; satisfaction with mental health; extent to which mental health interferes in daily life; Single question	Varied (ranges from current assessment to last 12 months)	~	~		×	◯			23 of 37

Note: ✓ indicates that the evidence shows this tool performs well on this dimension; ~ indicates that the evidence shows this tool performs only fairly; × indicates that the evidence shows this tool performs poorly; and ◯ indicates that evidence is limited or missing. If a cell is blank, this means that no research on this tool / topic combination was reviewed for this publication. * Refers to the fact that Germany included the longer SF-36 (rather than the shorter SF-12) in its 1998 German National Health Interview and Examination Survey, however the instrument will not be used in future due to licensing fees. Refer to Annex 2.A and Annex 2.B for more information about each tool. Country coverage refers to all OECD countries except Estonia, which did not participate in the questionnaire.
Source: Literature reviewed in this chapter; Responses to a questionnaire sent to national statistical offices in January 2022.

Statistical quality

A suitable measurement instrument for population mental health should perform well across a range of statistical qualities, including reliability, validity, ability to differentiate between different latent constructs, minimal non-response or refusals, comparability across groups and sensitivity to change. In addition, practical considerations surrounding a tool are important, such as keeping it short enough in length, with low redundancy between question items, so as to avoid respondent fatigue. These qualities interact with one another, meaning that in practice the goal is to balance the trade-offs of each in order to find a sensible solution. An instrument that performs well in one quality criterion – i.e. validity – may perform poorly in another – i.e. length of the questionnaire and/or non-response rates. Thus before choosing a metric, it is important for survey producers to weigh the costs and benefits of each approach to identify a tool suitable for their context.

How reliable are survey measures of mental health?

Measures of population mental health should produce consistent results when an individual is interviewed or assessed under a given set of circumstances. This concept, called reliability, is about ensuring that any changes detected in outcomes have a low likelihood of being due to problems with the tool itself – i.e. measurement error – and instead reflect actual underlying changes in the individual's mental health (Box 3.1).

Box 3.1. Statistical definitions: Reliability

Two important aspects of reliability are test-retest reliability and internal consistency reliability (OECD, 2013[1]; OECD, 2017[2]).

Test-retest reliability concerns a scale's stability over time. A respondent is re-interviewed or re-assessed after a period of time has passed, and their responses to a given questionnaire item are compared to one another. The expectation is that (assuming no change in the underlying state being measured) a reliable measure should lead to responses that are highly correlated with one another. There is no fixed rule for the length of time between the initial interview and follow-up: practice ranges from as short as 2-14 days to six months, depending on the assessment type (NHS Health Scotland, 2008[3]).

The test-retest criterion must be applied thoughtfully in the case of mental health measurement instruments, as mental health states (and particularly affective states) can fluctuate over short periods of time for a given individual. This means that measurement instruments addressing *specific symptoms* or *states* can be highly reliable yet still produce different results for the same individual over a period of days or weeks, as symptoms and experiences themselves wax and wane. In the context of measuring *population* mental health outcomes, then, test-retest reliability is particularly relevant for:

- Simple measures that concern whether an individual has ever been diagnosed with a mental health condition (where a good instrument should have a very high test-retest correlation)
- Establishing whether a short-form measure (or a measure being validated against a clinical diagnosis) is performing with the same test-retest accuracy as a long-form measure (or clinical diagnosis) when the two are administered to the same respondent, and/or
- Establishing the broad stability of symptom-based measurement scales over short time periods and across large samples - i.e. while the test-retest correlation of questions for a set of symptoms is unlikely to be perfect for a given individual (if symptoms themselves are not always

stable), day-to-day fluctuations in symptoms at the individual level can be expected to wash out across large samples to produce a similar distribution of scores over a short time period.[2]

Assessing test-retest reliability therefore indicates a trade-off between measures that are sufficiently stable, yet sensitive to change over time. An instrument that performs well on test-retest reliability may perform poorly on tests to measure sensitivity to change, which underscores the importance of looking at statistical quality measures holistically when making decisions as to which tools to implement.

Internal consistency reliability assesses the extent to which individual items within a survey tool are correlated to one another when those items aim to capture the same target construct. In the context of measuring population mental health, this might mean that, in a battery of items designed to measure depression and anxiety, the depression items correlate with one another, and the anxiety items correlate with one another (see also Box 3.3 for a discussion of factorial validity). The most widely used coefficient for internal consistency reliability is Cronbach's alpha, which is a function of the total number of question items, the covariance between pairs of individual items and the variance of the overall score.[1] Although there is not universal consensus, most researchers agree that a coefficient value between 0.7 and 0.9 is ideal (NHS Health Scotland, 2008[3]). Values below 0.7 may reflect the fact that items within the scale are not capturing the same underlying phenomenon (OECD, 2013[1]), while values above 0.9 may indicate that the scale has redundant items.

Notes:
1. The Cronbach coefficient alpha is commonly used in the literature to assess the internal consistency reliability of multi-item tools. The coefficient is calculated by multiplying the mean paired item covariance by the total number of items included in the scale and dividing this result by the sum of all elements in the variance-covariance matrix (OECD, 2013[1]). This results in a coefficient ranging from 0 (scale items are completely independent from one another, no covariance) to 1 (scale items overlap, complete covariance).
2. The definition of "a short time period" is subjective and can vary depending on circumstance. For example, although the period of a couple of days may be deemed an acceptably short period of time over which a test-retest assessment could be administered, if there were to be an extreme shock in the intervening days, either positive or negative, there would be good grounds to expect change in the underlying distribution. Frequent data collection on mental health during the COVID-19 pandemic illustrated the volatile nature of many affect-based measures, with large spikes coinciding with the introduction / easing of confinement policies.

The performance of screening tools on measures of reliability varies across tools and the outcomes they measure. There are mixed findings for general measures of psychological distress. The General Health Questionnaire (GHQ-12) as well as the Short Form-36 (SF-36) and its shorter sub-component, the Mental Health Inventory (MHI-5), have been shown to have good reliability (Schmitz, Kruse and Tress, 2001[4]; Ohno et al., 2017[5]; Elovanio et al., 2020[6]; Strand et al., 2003[7]); however, while the longer Kessler (K10) has been shown to be internally consistent, the test-retest reliability of the shorter Kessler (K6) tool has not been assessed in any studies (El-Den et al., 2018[8]; Easton et al., 2017[9]).

Conversely, screening tools for specific mental conditions – especially depression – are the most studied, and they have been shown to be reliable in terms of both test-retest reliability and internal consistency reliability. A meta-analysis of 55 different screening tools for depression found the Patient Health Questionnaire (PHQ-9) to be the most evaluated tool, with a number of studies concluding that both it and the PHQ-8 (a shorter version with the final question on suicidal ideation removed) have high reliability and validity (El-Den et al., 2018[8]). The same report, however, found that the shorter Patient Health Questionnaire-2 (PHQ-2) lacked consistent data on validity and reliability: among the six reports that evaluated the PHQ-2, only one reported on its internal consistency or test-retest reliability (El-Den et al., 2018[8]), which led the authors to caution that the reliability of the PHQ-2 cannot be confirmed with available data. The Center for Epidemiological Studies Depression Scale (CES-D), although less studied than the PHQ, has also been found to have good reliability, on both metrics (Ohno et al., 2017[5]). Among anxiety tools, the Generalised Anxiety Disorder screeners (both the longer GAD-7 and shorter GAD-2) have been found to be reliable, with good test-retest and internal consistency reliability (Ahn, Kim and Choi, 2019[10]; Spitzer et al., 2006[11]).

A study of the Patient Health Questionnaire-4 (PHQ-4), which combines the PHQ-2 and GAD-2 to generate a composite measure of both depression and anxiety, found lower, yet still acceptable reliability (Cronbach's alpha > 0.80 for both sub-scales) (Kroenke et al., 2009[12]). Another study of the PHQ-4 found lower item-intercorrelations but deemed the reliability to be acceptable given the short length of the scales (Löwe et al., 2010[13]).[3] Because Cronbach's alpha is in part a function of the total item length (refer to Box 3.1), shorter scales will perform worse on tests of internal consistency by construction. However shorter measures, with less redundancy between question items, are often preferred by survey creators, as they entail a lower burden for respondents.

Composite scales capturing aspects of positive mental health have also been found to be reliable**.** A study of the 14-question Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) tool found it to have high test-retest reliability (0.83 at one week) and a high Cronbach's alpha (around 0.9) (Tennant et al., 2007[14]; NHS Health Scotland, 2016[15]). The authors cautioned, though, that the high Cronbach's alpha suggests some redundancy in the scale items, a concern that led to the development of the shorter seven-item version (SWEMWBS) (Tennant et al., 2007[14]; NHS Health Scotland, 2016[15]). Multiple studies of WEMWBS and SWEMWBS found them both to have strong test-retest reliability (Stewart-Brown, 2021[16]; Shah et al., 2021[17]). The World Health Organization-5 (WHO-5) composite scale has also been tested for reliability in a variety of settings (Dadfar et al., 2018[18]; Garland et al., 2018[19]). Similarly, the MHC-SF has been found to have high internal reliability, though its test-retest reliability is only moderate (Lamers et al., 2011[20]).

Fewer studies have investigated the reliability of general self-reported indicators of mental health status; however, evidence from the United States suggests that these measures have acceptable test-retest reliability. The health-related quality-of-life tool used by the United States Centers for Disease Control, the Behavioral Risk Factor Surveillance System (BRFSS) survey, measures perceived health by combining physical and mental health. A study in the state of Missouri found that the shorter version of the tool, with four items, has acceptable test-retest reliability and strong internal validity, although reliability was lower among older adults (Moriarty, Zack and Kobau, 2003[21]).

Box 3.2. Key messages: Reliability

- Most mental health screening tools, including both surveys that identify specific mental disorders and those that identify positive mental health, have been found to have strong reliability, as measured through both test-retest and internal consistency measures.
- Test-retest reliability must be considered in tandem with a measure's sensitivity to change over time, rather than blindly applied as a quality criterion.
- There is strong evidence for the reliability of screening tools (especially those focusing on depression) and, to a somewhat lesser extent, positive mental health composite scales. However, fewer studies have been done to assess the reliability of general self-reported indicators of mental health status; more research is needed in this area.

How well does the tool measure the targeted outcome?

In addition to being reliable, a good measurement instrument must be valid, i.e. the measures provided by the tool should accurately reflect the underlying concept. For indicators that are more objective, validity can be assessed by comparing the self-reported measure against an objective measure of the same construct. For example, respondents' self-reported earnings could in theory be cross-checked with their tax returns, or pay slips, to ascertain whether their response was reported accurately. Of course there are practical reasons that prevent this from being done systematically, but this illustrates that there are ways of assessing the validity of self-reported earnings data. Conversely, it is not possible to ascertain the "objective truth" of a subjective indicator, such as subjective well-being, trust or indeed mental health. This

does not mean that validity cannot be assessed: OECD measurement guidelines use the concepts of face validity, convergent validity and construct validity to assess the validity of subjective indicators (OECD, 2013[1]; OECD, 2017[2]) (Box 3.3).

Unlike many subjective indicators, the bulk of screening tools to assess mental health have been validated against diagnostic interviews for common mental disorders, which provide a rigorous assessment of their accuracy and real-world meaning. The most common diagnostic interview against which mental health screening tools are validated is the World Health Organization's Composite International Diagnostic Interview (WHO-CIDI), which was designed for use in epidemiological studies as well as for clinical and research purposes (see Chapter 2 for more details). This tool allows to measure the prevalence of mental disorders, the severity of these disorders, their impact on home management, work-life balance, relationships and social life, as well as mental health service and medications use. Although the CIDI is widely accepted as a gold standard against which mental health survey items should be assessed, it is not immune to criticisms and validity concerns (Box 3.4).

Box 3.3. Statistical Definitions: Validity

Validity is more difficult to ascertain than reliability, especially for subjective data for which an objective truth is unknowable, and which typically cannot be compared to an equivalent objective measure. Three ways of assessing validity for subjective measures include face validity, convergent validity and construct validity.

Face validity evaluates whether the indicator makes intuitive sense to the respondent and to (potential) data users. One way to indirectly measure face validity is through non-responses. High levels of non-response may indicate that respondents do not understand or see the relevance or usefulness of the question. In the case of mental health, high levels of non-response may also reflect a degree of discomfort with the topic due to stigma and bias, rather than lack of face validity. (An extended discussion of non-response and mental health measures appears later in this chapter.) Cognitive interviewing can also be used.

Convergent validity is assessed by how well the indicator correlates to other proxies of the same underlying outcome. Using mental health tools as an example, were a researcher to introduce a new tool to assess anxiety, s/he could test its convergent validity by comparing it to pre-existing screening tools for data on anxiety, diagnosis or mental health service use, self-reported assessments of anxiety level, and/or bio-physical markers of stress and anxiety (heart rate, blood pressure, neuroimaging, etc.).

Construct validity is the extent to which the indicator performs in accordance with existing theory or literature. For example, research shows that mental health and physical health are correlated with one another and co-move. Therefore, if a new mental health tool showed little correlation with physical health, or if changes in mental health as measured by this tool did not reflect any changes in physical health, the scale would be suspected of having low construct validity. The growing literature on the social determinants of health can also be leveraged to assess construct validity, in a similar way.

In addition to the three aspects of validity mentioned above, clinical validations of mental health survey items often refer to three additional assessments: criterion validity, factorial validity and cross-group validity.

Criterion validity exists only when there is a gold standard against which an item can be compared. In the case of mental health, this gold standard is typically a structured interview (e.g. the CIDI, refer to Annex 2.B). Criterion validity assesses the psychometric properties of a measure, i.e. how it compares to the gold standard. A measure is said to be sensitive if it can accurately identify a "true positive" (i.e. how often the survey accurately identifies someone at risk of, say, depression); it is specific if it can accurately identify a "true negative" (i.e. it accurately identifies someone as *not* at risk for depression).

In order to establish diagnostic accuracy, sensitivity and specificity are plotted in a receiver operating characteristic (ROC) curve at various thresholds. The area under the curve (AUC) can then be used to assess the diagnostic performance of the screening tool in comparison to the gold standard.[1]

Factorial validity assesses whether a multi-item survey tool is measuring one, or several, underlying concepts. In almost all cases, unidimensionality is desired if only a single construct is being assessed; this provides assurance that the mental health tool is measuring, for example, depression, anxiety or latent well-being. However, if a scale is assessing multiple dimensions of mental health, then multidimensionality is desired. For example, factor assessments for the PHQ-4, which measures depression and anxiety, indeed identify two latent factors (Löwe et al., 2010[13]). Factorial validity is commonly assessed using either confirmatory factor analysis (CFA) or exploratory factor analysis (EFA). In the former, researchers test a hypothesis that the relationship between an observed variable (e.g. respondents' answers to the PHQ-8 tool) and an underlying latent construct (e.g. depression) fits a given model. That is, using CFA, researchers test the hypothesis that an observed dataset has a given number of underlying latent factors. Using EFA, researchers do not impose a theoretical model and instead work backwards to uncover the underlying factor structure (Suhr, 2006[22]).

Cross-group validity, or cultural validity, refers to the extent to which a measure is applicable across different population groups. There are a range of ways that cross-group validity can bias mental health outcome measures, including through cultural factors affecting the way in which symptoms are expressed, clinical bias (either implicit or explicit), language limitations of the respondent (if the tool is being implemented in a language other than their mother tongue) and differences in response behaviour (e.g. greater likelihood to choose midpoint values on Likert scales rather than extreme values) (Leong, Priscilla Lui and Kalibatseva, 2019[23]). Cross-group validity is best ensured by validating a survey tool in the requisite population, rather than applying it blindly.

Notes:

1. A receiver operating characteristic curve provides a visualisation of diagnostic ability by plotting the true positive rate against the true negative rate. The curve can be used to determine the optimal cut-off point, which minimises both Type 1 (false positive) and Type II (false negative) errors. ROC analysis is used in determining the threshold cut-off scores, which are discussed later in this chapter. For more information on ROC and its use in clinical psychology, refer to (Pintea and Moldovan, 2009[24]) and (Streiner and Cairney, 2007[25]).

To assess the validity of screening tools, researchers typically implement a study in which respondents both answer the self-reported scale and participate in a structured CIDI interview, with their responses to both then compared. A screening tool with high sensitivity and specificity is said to have high criterion validity. Although criterion validity ensures that screening tools are designed to mirror diagnostic outcomes from the CIDI, screening tools by design estimate higher prevalence rates for specific mental disorders (see Box 2.1). Convergent validity is assessed by comparing different screening tools against one another to see whether a new tool for measuring, say, depression, performs similarly to existing measures for depression. This approach is often used when testing shortened versions of screening tools, to see whether the truncated survey performs as well as its longer, more in-depth, predecessor. The majority of screening tools described in this chapter have been validated against diagnostic interviews for common mental disorders and have reported good psychometric properties (high sensitivity and specificity) across age groups, gender and socio-economic status (Gill et al., 2007[26]; O'Connor and Parslow, 2010[27]; Huang et al., 2006[28]) (Box 3.3).

Box 3.4. Validity of structured interviews

One important caveat to using structured interviews to validate screening tools is that it presupposes the structured interviews to be an accurate measure of "true" underlying mental health. This issue is raised in two different contexts: (1) most screening tools used in OECD countries were validated against the fourth version of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV), published in 1994, which is now outdated, rather than against the DSM-5, published in 2013; and (2) the extent to which the DSM itself provides accurate diagnostic data cross-culturally.

The first concern relating to the validity of structured interviews has to do with the fact that none of the screening tools commonly in use have been validated against the newer DSM version (Statistics Canada, 2021[29]). Yet, in total, there are 464 differences between the DSM-IV and DSM-5. Broadly speaking, the DSM-5 includes fewer diagnostic categories, as many previously separate disorders share a number of features or symptoms. In addition, greater effort was made to separate an individual's functioning status from their diagnosis. One area that could have an impact moving forward is the lowering of the diagnostic threshold for generalised anxiety disorder – a move that has been criticised by some psychiatrists for pathologising what had previously been considered quotidian worries (Murphy and Hallahan, 2016[30]).

In sum, even though there are always changes between DSM updates that include the restructuring of diagnostic categories and the updating of some diagnosis criteria, there is by design a degree of continuity between different DSM versions, and most changes are minor. Regardless, in order to be up to date with most recent clinical practice, instruments like the CIDI would benefit from an update.

On the second point, there are concerns about the applicability of these diagnostic validations to non-US regions and population groups, which at the very least would require validation studies to be conducted in different local contexts. Beyond this, validating mental health screening tools in more geographically diverse clinical settings may be insufficient if the clinical diagnoses underpinning the validation are themselves flawed. Haroz and colleagues investigated the extent of this cross-cultural bias by reviewing 138 qualitative studies of depression reflecting 77 different nationalities and ethnicities (Haroz et al., 2017[31]). They found that only 7 of the 15 most frequently mentioned features of depression across non-Western populations reflect the DSM-5 diagnosis of Major Depressive Disorder. DSM-specified diagnostic features including "problems with concentration" and "psychomotor agitation or slowing" did not appear frequently, while features including "social isolation or loneliness", "crying", "anger" and "general pain" – none of which are included as diagnostic criteria – did. Some features arose more frequently in certain regions: "worry" in South and Southeast Asia, and "thinking too much" in Southeast Asia and sub-Saharan Africa. This implies that the close alignment of the PHQ-9 or the GAD-7 with the DSM criteria could in theory limit detection of the underlying targeted construct (i.e. depression, anxiety) relative to longer or more comprehensive screening tools and/or structured interviews (Ali, Ryan and De Silva, 2016[32]; Sunderland et al., 2019[33]).

Although criticisms of DSM criteria do exist, the DSM still remains the most useful tool for enabling cross-country comparative data on mental health outcomes. While improvements could be made, the DSM includes considerations of cultural validity in its drafting, which are updated in each subsequent iteration.

Moving beyond clinical psychology, a few OECD countries have expanded their definition of mental health to encompass a wider range of viewpoints, beyond the traditional ones rooted in a Western perspective. In New Zealand, for example, the Government Inquiry into Mental Health and Addiction includes a Māori perspective of mental health (New Zealand Government, 2018[34]). In a similar vein,

the Swedish government will solicit feedback from the Sami parliament when drafting its upcoming strategy on mental health (Public Health Agency of Sweden, 2022[35]).

Across measures of general psychological distress, the Kessler and MHI-5 scales have stronger criterion validity than the GHQ-12. Studies have found that the K10 and K6 scales have strong psychometric properties (encompassing both reliability and validity) and better overall discriminatory power than the GHQ-12 in detecting depressive and anxiety disorders (Furukawa et al., 2003[36]; Cornelius et al., 2013[37]). The mental health component of the SF-12 tool is also better able to discriminate between those with and those without specific mental health conditions, as compared to the GHQ-12 (Gill et al., 2007[26]).While the MHI-5 tool has been found to be just as valid as the longer MHI-18 and GHQ-30 to assess a number of mental health conditions, including major depression and anxiety disorders, it performed less well than the MHI-18 for the full range of affective disorders (Berwick et al., 1991[38]). While the MHI-5 was designed as a general tool, it has been proven effective to identify a specific risk for depression and/or anxiety (Yamazaki, Fukuhara and Green, 2005[39]; Rivera-Riquelme, Piqueras and Cuijpers, 2019[40]).

A recent meta-analysis of the sensitivity and specificity of instruments used to diagnose and grade the severity of depression reported that, on average, the PHQ-9 demonstrated the highest sensitivity and specificity relative to other screening tools, including the CES-D (Pettersson et al., 2015[41]). A different version of the PHQ-8 has been used in the CDC Behavioral Risk Factor Surveillance System (BRFSS) survey. This measure, referred to as the PHQ-8 days, asks respondents how many days over the past four weeks they have experienced each of the eight depressive symptoms that make up the PHQ-8. This yields a scale ranging from 0-112 and can provide a look at depression risk that is more granular – better identifying individuals who may be at risk for mild depression but currently have higher levels of mental well-being – and also more sensitive to change (Dhingra et al., 2011[42]). The PHQ-2 has been assessed for its internal consistency, construct validity and correlation convergent validity; however, a meta-analysis did not find evidence of studies of criterion validity (El-Den et al., 2018[8]). Another overview cites evidence for the PHQ-2 as having good criterion validity for specific populations such as older adults, pregnant or post-partum women, and patients with specific conditions such as coronary heart disease or HIV/AIDS (Löwe et al., 2010[13]).

While self-report scales for depressive symptoms tend to be well validated, scales for anxiety disorders have been found to be somewhat less sensitive and specific in clinical populations. Research suggests this may be because different types of anxiety disorders have more heterogeneous symptoms than depressive disorders (Rose and Devine, 2014[43]). Despite this, both the GAD-7 and GAD-2 have been validated in a number of studies. The GAD-7 was designed to provide a brief clinical measure of generalised anxiety disorder, and its validation exercise found it to have good validity (criterion, construct, factorial, etc.). Furthermore, factorial validity assessments of the GAD-7 and PHQ-8 found that, despite a high correlation between the anxiety and depression scales (0.75), the two scales are complementary and not duplicative; more than half of patients with high levels of anxiety did not also have high levels of depression (Spitzer et al., 2006[11]). The high correlations of the GAD-7 with two other anxiety scales indicated good convergent validity (Kroenke et al., 2007[44]; Spitzer et al., 2006[11]).[4] Both the GAD-7 and the shorter GAD-2 perform well in detecting all four major forms of anxiety disorders: generalised anxiety disorder, panic disorder, social anxiety disorder and post-traumatic stress disorder (Kroenke et al., 2007[44]).

The PHQ-4 has been found to be a valid tool for measuring the combined presence of risks for both depression and anxiety. As noted above, its component parts – the PHQ-2 and GAD-2 – have been validated against diagnostic criterion standard interviews (with caveats to the broader applicability of PHQ-2 criterion validity, as mentioned above). Studies have shown that PHQ-4 scores are associated with the SF-20 functional status scale and health information such as disability days used, etc., providing evidence for convergent and construct validity. Furthermore, factorial analysis has found that the PHQ-4 has a two-

dimensional structure with two discrete factors, picking up on both depression and anxiety disorders (Löwe et al., 2010[13]).

Composite scales capturing aspects of positive mental health have also been found to have good validity. WEMWBS was found to have good criterion and convergent validity, being highly correlated with other scales that capture positive affect. WEMWBS and the WHO-5 are, unsurprisingly, highly correlated with one another (correlation coefficient of 0.77) (NHS Health Scotland, 2016[15]), with WEMWBS being slightly less correlated with other measures of mental health that had a stronger focus on physical health or psychological distress (including the GHQ-12). Another study on WEMWBS found that the shorter version of the screening test was highly correlated with the longer version, making it an efficient and quicker alternative to the longer 14-question version (Stewart-Brown et al., 2009[45]). Despite its length, Rasch analysis has found that WEMWBS is unidimensional with one underlying factor (Stewart-Brown, 2021[16]).[5] Multiple studies have shown the MHC-SF has good convergent validity (Guo et al., 2015[46]; Petrillo et al., 2015[47]; Lamers et al., 2011[20]), however cognitive interviews in Denmark found that it had poor face validity, especially for questions on the social subscale (Santini et al., 2020[48]).

Although designed as measures of positive mental health, both WEMWBS and the WHO-5 have been shown to be effective screeners for depression and/or anxiety. A study found the WHO-5 to have high sensitivity, but low specificity, in identifying patients with depression in a clinical setting (Topp et al., 2015[49]). A study of SWEMWBS found it to be relatively highly correlated with the PHQ-9 (rho = 0.6-0.8) and the GAD-7 (rho = 0.6-0.7), suggesting that it is an acceptable tool for measuring common mental disorders (CMD); however, other tools may be more sensitive in identifying and distinguishing between individuals with worse levels of mental health (Shah et al., 2021[17]). A study comparing WEMWBS to the GHQ-12, through multidimensional item-response theory, found that both tools appear to measure the same underlying construct (Böhnke and Croudace, 2016[50]).

Self-reported mental health (SRMH) indicators have been compared to validated clinical measures of mental health and have been shown to be related to, though distinct from, other mental health scales. SRMH is correlated with the Kessler scales, the PHQ and the mental health component of the SF-12 and is often used in the validation process of other mental health screening tools as a test for convergent validity. Furthermore, SRMH is associated with poor physical health and an increased use of health services. Although related, research has shown that correlations between SRMH and screening tools are moderate, suggesting that they are capturing slightly different phenomena (Ahmad et al., 2014[51]). The authors note that further research is needed but suggest that findings from longitudinal studies of self-reported *physical* health could shed some light. SRMH measures were shown to be stronger predictors of mortality, morbidity and service use than other indicators, and that SRMH may be capturing mental health problems that do not yet manifest in screening tools (Ahmad et al., 2014[51]). Conversely, health-related quality of life (HRQoL) – which measures both physical and mental health – has been found to have construct and criterion validity that is good and comparable to the SF-36 scale (Moriarty, Zack and Kobau, 2003[21]).

Studies of mental health screening tools have yielded conflicting evidence as to whether single-item mental health questions are sufficiently valid. A study assessing the comparative performance of the MHI-5 and MHI-18 (which concluded in favour of the shorter version) found that even a single question – “how often were you feeling downhearted and blue?” – performed as well as the MHI-5, MHI-18 and GHQ-30 at detecting major depression (Berwick et al., 1991[38]). However, studies assessing ultra-short screening tools found that even two questions perform significantly better at screening for depression than does a single question (Löwe et al., 2010[13]). Conversely, the Australian Taking the Pulse of the Nation (TPPN) survey, administered throughout the COVID-19 pandemic, found that the psychometric properties of its single-item mental health measure compared favourably to the K6: the items were highly correlated (rho = 0.82), and the single-item measure had high sensitivity for psychological distress (Botha, Butterworth and Wilkins, 2021[52]).

Box 3.5. Key messages: Validity

- All of the mental health screening tools commonly used by OECD member states have been validated in clinical settings and found to have strong convergent, construct and criterion validity.
- Composite scales for positive mental health have also been found to have strong psychometric properties, and they have proven effective as screeners for specific mental health conditions such as depression and/or anxiety.
- Criterion validity is assessed by the survey tool's performance in comparison to a clinical diagnostic interview gold standard; however, this presupposes the validity of clinical diagnoses, which may not hold in all contexts.

What do non-response rates tell us about stigma? How does this affect the comparability of mental health data across groups?

The stigma associated with mental illness may lead to misreporting – and under-reporting – of one's mental health conditions (Hinshaw and Stier, 2008[53]). Low levels of mental health literacy can also lead to under-reporting, with individuals not recognising their own experienced symptoms as representative of an underlying condition (Tambling, D'Aniello and Russell, 2021[54]; Dunn et al., 2009[55]; Coles and Coleman, 2010[56]).[6] Feelings of stigma towards mental health conditions remain important in all OECD countries, with large differences between them. A survey conducted in 2019 found that, in 19 OECD countries, 40% of respondents did *not* agree with the statement that mental illness is just like any other illness, and a quarter agreed that anyone with a history of mental disorders should be prevented from running for public office (OECD, 2021[57]). Because of stigma, respondents may either conceal their true conditions when answering mental health surveys or may choose not to participate in the first place. When administering surveys on sensitive subjects, providing clear assurances of data confidentiality and ensuring that the interview is conducted in a private place, out of hearing of family members, minimise the likelihood of respondent refusal (Singer, Von Thurn and Miller, 1995[58]; Krumpal, 2013[59]).[7]

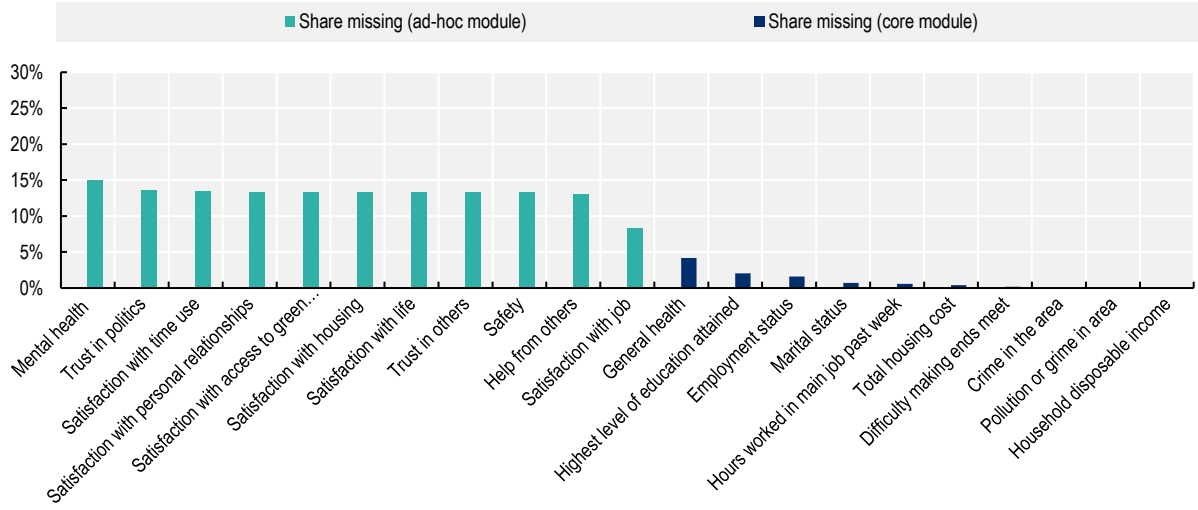
Evidence shows that those experiencing psychological distress or a specific mental disorder are more likely to refuse to participate in a survey; this non-response bias then leads to underestimates of the overall prevalence of mental ill-health (de Graaf et al., 2000[60]; Eaton et al., 1992[61]; Mostafa et al., 2021[62]). A recent study, which compared the effect of psychological distress on a number of economic transitions (e.g. falling into unemployment), using both the GHQ-12 score and a version of it adjusted for misreporting behaviour scores, showed that the original version of the GHQ-12 score underestimated the effect of psychological distress on transitions into better-paid jobs and higher educational attainment (Brown et al., 2018[63]). Thus, misreporting of symptoms of psychological distress can lead to biased and inconsistent estimates. However, not all studies come to the same conclusion: the US National Comorbidity Survey Replication (NCS-R) study found no evidence of non-response leading to underestimates of disorder prevalence (Kessler et al., 2004[64]).

Evidence from the 2013 European Union Statistics on Income and Living Standards (EU-SILC) survey shows that non-response rates for mental health questions are high (15%), but still comparable to those for other subjective variables (e.g. 13% for trust in politics, 8% for satisfaction with one's job (Figure 3.1 Panel A). High non-response rates for mental health (as measured through the MHI-5) may partly reflect the way in which the EU-SILC survey is implemented. Each year, an ad-hoc module featuring additional questions on a specific topic is implemented in addition to the core module (in 2013 this module focused on well-being), implying that some respondents may have problems in answering questions that were not asked in previous waves.

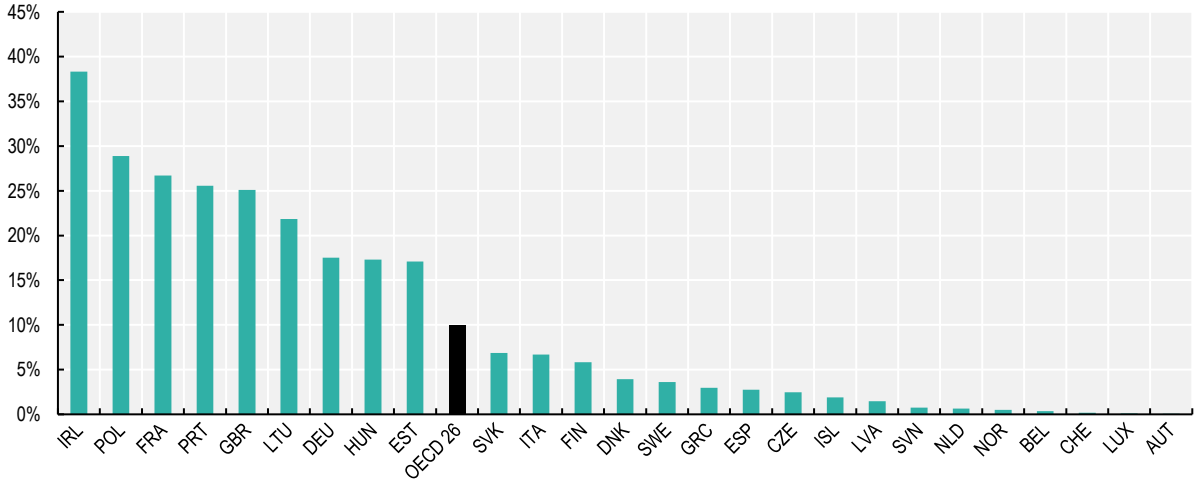
Missing and non-response rates for mental health variables vary widely between countries (Figure 3.1, Panel B). In the 2013 EU-SILC survey, missing rates for the mental health module were higher than 20% in Ireland, Poland, France, Portugal, the United Kingdom and Lithuania, but were below 1% in Norway, Belgium, Switzerland, Luxembourg and Austria.

Figure 3.1. Non-response rates are higher for mental health questions than they are for other variables, and vary substantially across countries

Panel A: Share of non-response rates for different types of variables, European OECD 26, 2013



Panel B: Share of non-response rates for mental health questions (MHI-5), European OECD 26, 2013



Note: This figure only includes individuals who have agreed to participate in the survey, and subsequently choose not to answer individual question items; it does not consider those who refuse to participate in the full survey. A respondent is deemed to be missing mental health data if they refused, or replied "do not know", to at least four of the five individual items on the MHI-5. Refer to Annex 2.B for more information about specific tools.

Source: OECD calculations based on the *European Union Statistics on Income and Living Conditions (EU-SILC)* (n.d.[65]) (database), https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions.

StatLink https://stat.link/s3v124

Table 3.2 shows some suggestive evidence that differences in non-response rates by country could be related to levels of stigma; in nine European OECD countries, the prevalence of any depressive disorder (as measured by the PHQ-8) is inversely correlated with the prevalence of mental health stigma, as measured by the share of the population who agree that people with a history of mental health problems should be excluded from running for office. Therefore, in countries with more stigma, the prevalence of depressive disorder risk is also lower – perhaps because of reluctance to report.[8]

Table 3.2. The relationship between stigma and prevalence is complex, but in some instances, stigma may lead to lower reported prevalence of mental disorders

Correlations between indicators of stigma towards mental health and prevalence of mental health conditions, nine European OECD countries

	Exclude from office if mental health history	Seeking treatment shows strength	Mental health like any other illness	Prevalence of psychological distress (MHI-5)	Prevalence of any depressive disorder (PHQ-8)	Share missing psychological distress responses (MHI-5)
Exclude from office if mental health history	1					
Seeking treatment shows strength	0.11	1				
Mental health like any other illness	0.11	0.56	1			
Prevalence of psychological distress (MHI-5)	0.14	-0.44	-0.47	1		
Prevalence of any depressive disorder (PHQ-8)	-0.83***	-0.19	-0.06	0.13	1	
Share missing psychological distress (MHI-5) responses	0.17	-0.6	0.11	-0.001	0.03	1

Note: Table displays listwise correlations. The three stigma questions ask respondents to indicate the extent to which they agree with the statement. For the first, agreement entails stigma; for the second two, agreement entails the absence of stigma. For details on the MHI-5 and PHQ-8 measures, see Annex 2.B. * Indicates that Pearson's correlation coefficients are significant at the p<0.10 level, ** at the p<0.05 level, and *** at the p<0.01 level. N = 9 countries.

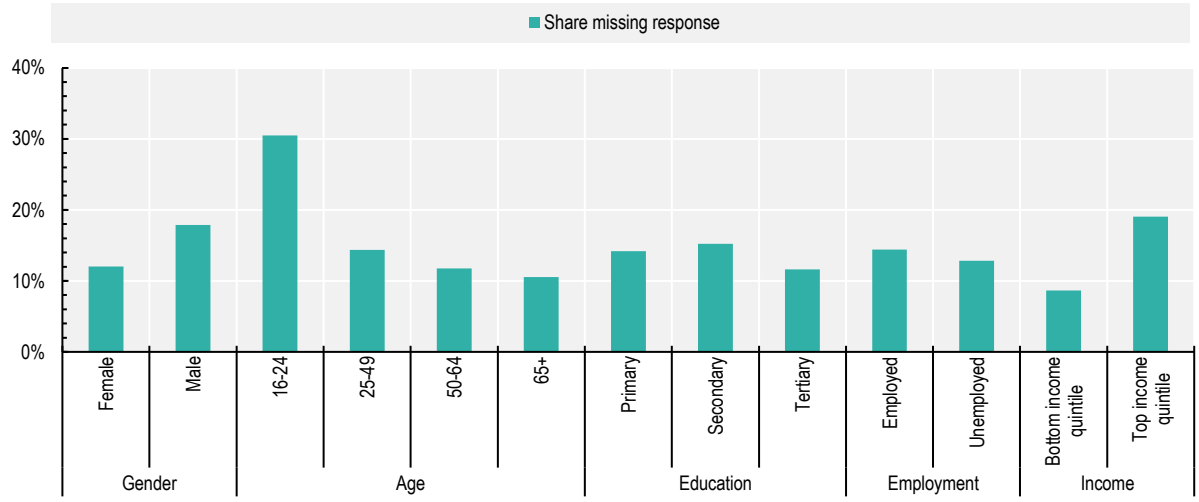
Source: Stigma data originally come from an Ipsos survey, as published in OECD (2021[57]), *Fitter Minds, Fitter Jobs: From Awareness to Change in Integrated Mental Health, Skills and Work Policies*, Mental Health and Work, OECD Publishing, Paris, https://dx.doi.org/10.1787/a0815d0f-en; MHI-5 data come from OECD calculations based on the 2018 *European Union Statistics on Income and Living Conditions* (EU-SILC) (n.d.[65]) (database), https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions; PHQ-8 come from OECD calculations based on *European Health Interview Survey* (EHIS) wave 3 data (n.d.[66]) (database), https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:European_health_interview_survey_(EHIS).

In order to understand what bias is introduced by non-response rates, it is important to understand the profile of those who are choosing not to respond to mental health questions. Figure 3.2 shows these shares for a number of socio-demographic groups – gender, age cohort, education level, labour market status and income quintile. Panel A displays non-responses to mental health questions for 26 European OECD countries, while Panel B shows those for the United Kingdom. For both data sources, women, those with higher levels of education, and older age cohorts are more likely to answer mental health questions, while men, young people and those with lower levels of education are more likely to not respond. These results are in line with a report describing stigma towards mental health in Sweden: women were found to be more likely than men to have positive feelings towards those with mental health conditions, while young people

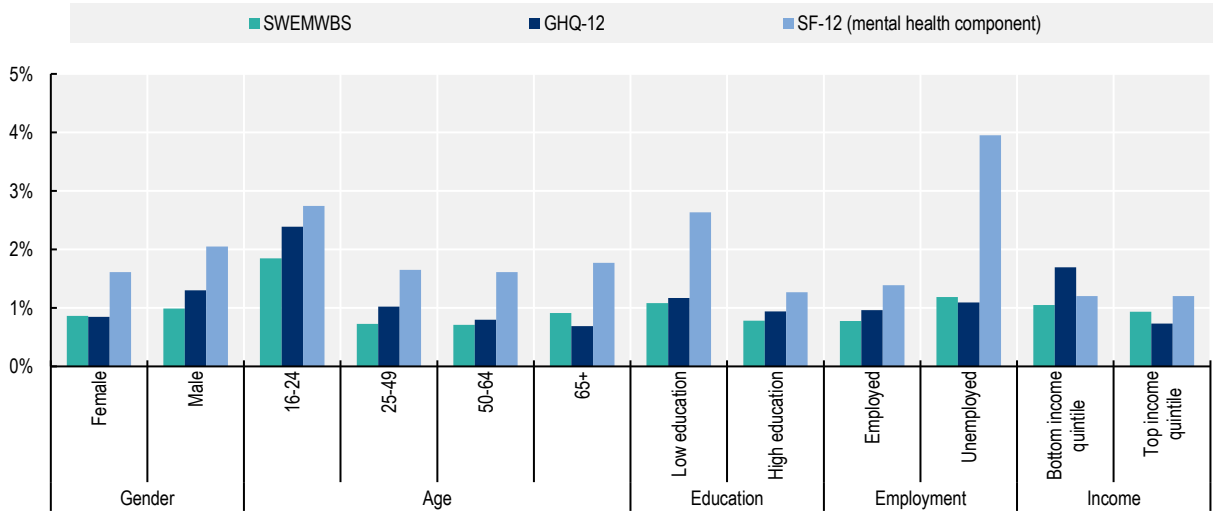
were more likely than older people to report that it would be difficult to talk about their own mental illness with someone else (Folkhälsomyndigheten, 2022[67]). In European countries, there is a clear difference by income – those in the top income quintile are less likely to respond – however, this pattern does not hold for the United Kingdom. A study on non-response rates in longitudinal health surveys among the elderly in Australia found that those with lower occupational status and less education were less likely to participate (Jacomb et al., 2002[68]); however, neither risk for depression nor anxiety influenced refusal rates. The Netherlands Mental Health Survey and Incidence Study-2 (MENESIS-2) found higher non-response rates among young adults, leading to under-reporting of specific mental disorders among this population (de Graaf, Have and Van Dorsselaer, 2010[69]).

Figure 3.2. Young people, men and those with lower levels of education are less likely to respond to mental health questions

Panel A: Share of non-response rates for mental health questions (MHI-5) across different socio-demographic groups, European OECD 26, 2013



Panel B: Share of non-response rates for mental health questions (SWEWMWBS, GHQ-12 and SF-12 mental health component), United Kingdom, 2019



Note: Refer to Annex 2.B for more information about specific tools.

Source: Panel A: OECD calculations based on the 2013 *European Union Statistics on Income and Living Conditions (EU-SILC)* (n.d.[65]), (database), https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions; Panel B: OECD calculations based on University of Essex, Institute for Social and Economic Research (2022[70]), *Understanding Society: Waves 1-11, 2009-2020 and Harmonised BHPS: Waves 1-18, 1991-2009* (database). 15th Edition. UK Data Service. SN: 6614, http://doi.org/10.5255/UKDA-SN-6614-16, from wave 10 only (Jan 2018 – May 2020).

StatLink https://stat.link/1im7az

Data on previous diagnoses for, or experienced symptoms of, specific mental disorders are likely to under-report population prevalence due to a combination of reticence to disclose personal medical history and inability to afford or access care (Hinshaw and Stier, 2008[53]). Furthermore, prevalence of mental ill-health based on these data is heavily influenced by the characteristics of health care systems in different countries and regions, including their ability to treat and diagnose a wide range of patients. For example, data predating the pandemic show that 67% of working-age adults who wanted mental health services reported difficulty in accessing treatment (OECD, 2021[71]). A survey in Canada compared the self-reported use of mental health services from the Canadian Community Health Survey with health service administrative data from the government of Quebec (*Régie de l'assurance maladie du Québec* – RAMQ), reporting significant discrepancies: 75% of mental health service users in the RAMQ registry did not report using services in the CCHS, with these disparities being highest for older people, those with lower levels of education and mothers of young children (Drapeau, Boyer and Diallo, 2011[72]). Another study for Australia examined the extent of under-reporting of mental illness by matching self-reported mental health information (self-report diagnosis and self-reports of prescription drug use) to administrative records of filled prescriptions for mental disorders; the researchers found that survey respondents are significantly more likely to under-report mental illnesses compared to other health conditions because of stigma (Bharadwaj, Pai and Suziedelyte, 2017[73]).

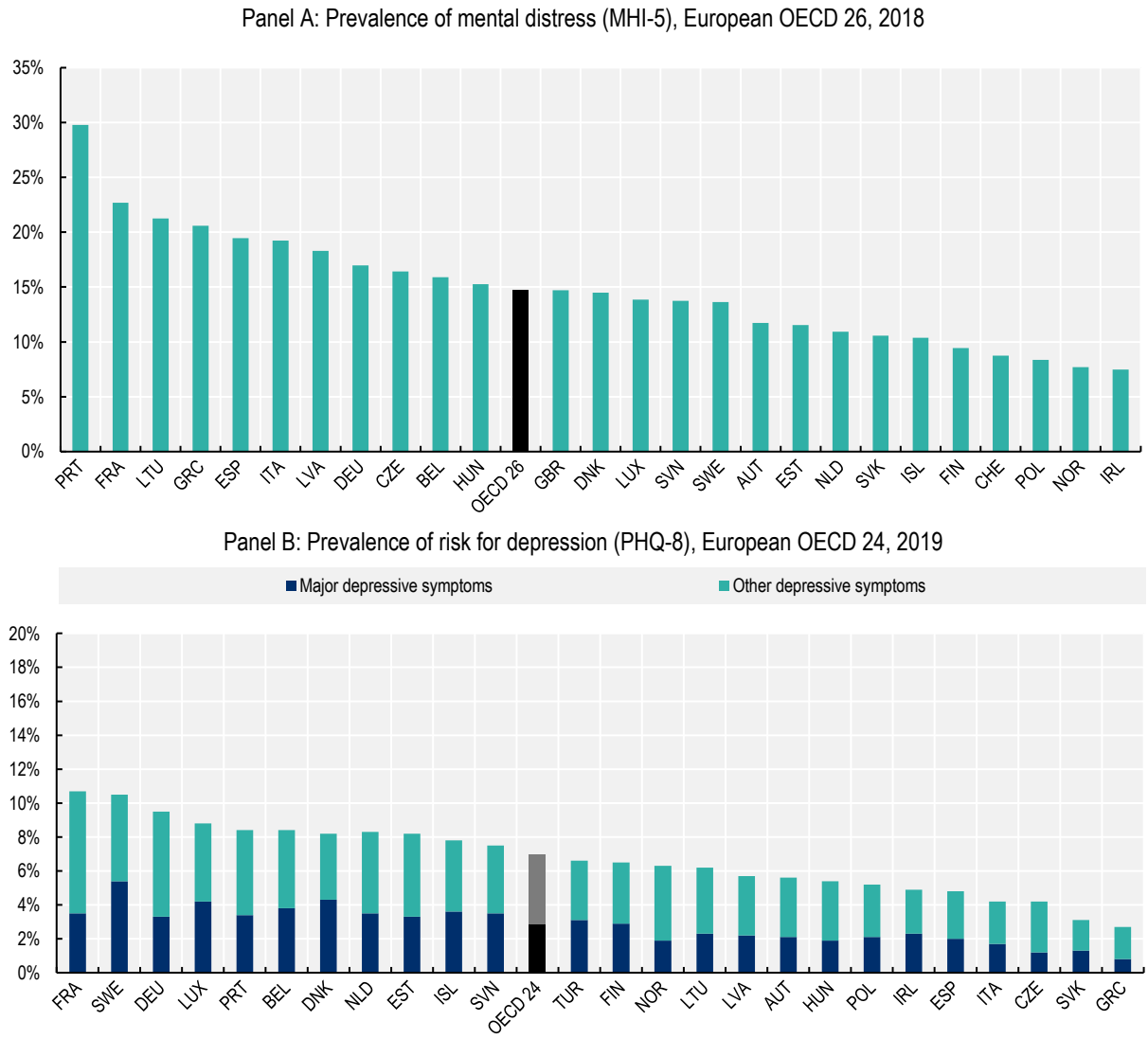
Box 3.6. Key messages: Non-response bias and missing values

- Those with worse underlying mental health may be more likely to refuse to participate in surveys, thereby understating the actual prevalence of mental ill-health; however, the evidence is not conclusive.
- There is conclusive evidence that self-reported data on previous diagnoses and experienced symptoms of specific mental ill-health conditions are significantly influenced by stigma and bias.
- Analysis from European OECD countries shows that younger people, men and those with lower levels of education are more likely to refuse to answer questions on mental health.

Are the reliability and validity of these measures consistent across cultures and socio-demographic groups?

Governments tasked with promoting population mental health need high-quality information to understand inequalities in mental health outcomes and whether national trends (either improvements or deteriorations) are masking differences within groups, so that they can target policy interventions to those who are most in need. For these reasons, a mental health indicator needs to be able to compare age cohorts, genders, race and ethnic groups, different education and income levels and other socio-economic markers. Ensuring comparability, however, is not straightforward. Cultural differences in perceptions of mental health may make some groups less likely to answer (or honestly answer) questions surrounding mental health. These challenges are true for both inter- and intra-country comparisons.[9]

Figure 3.3. Prevalence of psychological distress and depressive symptoms risk varies by as much as 100 percent across European OECD countries



Note: Panel A: risk for psychological distress is defined as having a score >= 52 on a scale from 0 (least distress) to 100 (most); Panel B: a respondent is deemed to be at risk for major depressive disorder if they answer "more than half the days" to either of the first two questions on the PHQ-8, and, in addition, if five or more of the eight items are reported as "more than half the days". They are at risk for "other depressive disorders" if they answer "more than half the days" to either of the first two questions on the PHQ-8, and in addition, a total of two to four of the eight items are reported as "more than half the days" (Eurostat, n.d.[74]). Refer to Annex 2.B for more information on individual screening tools.
Source: Panel A: OECD calculations based on the 2018 *European Union Statistics on Income and Living Conditions (EU-SILC)* (n.d.[65]), (database), https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions; Panel B: *European Health Interview Survey* (EHIS) wave 3 data (n.d.[66]) (database), https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:European_health_interview_survey_(EHIS).

StatLink https://stat.link/ocxvgt

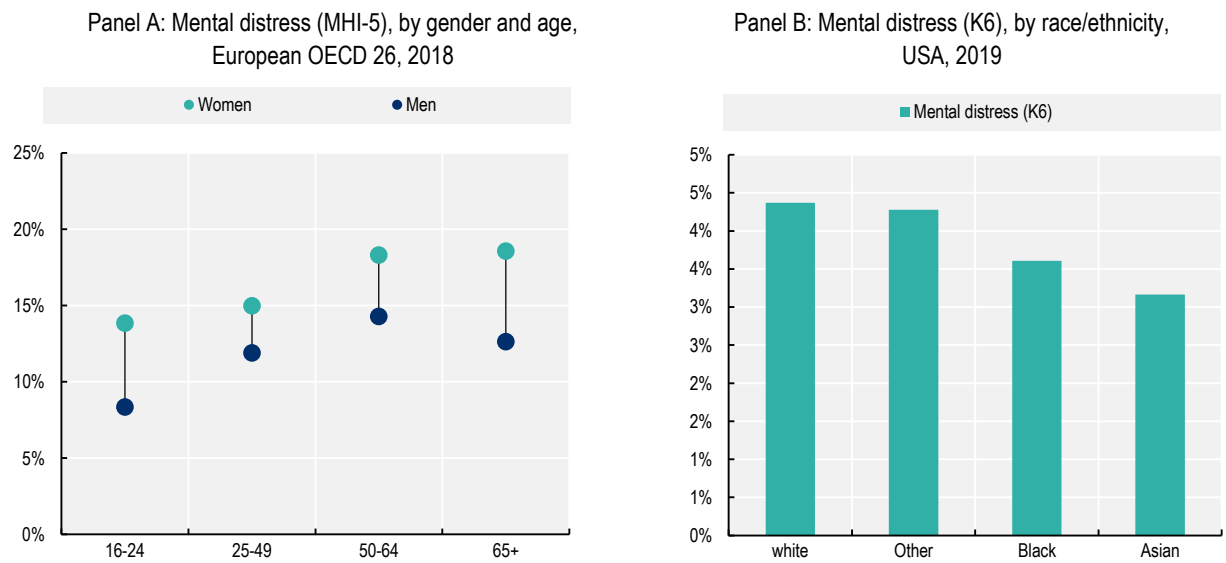
Data from European countries show large variations in the prevalence of psychological distress and depressive symptoms. The prevalence of psychological distress in Portugal, France and Lithuania is more than twice that of Ireland, Norway, Poland and Switzerland (Figure 3.3, Panel A). Similarly, the prevalence of depressive symptoms in France, Sweden and Germany is more than double that of Greece, the Slovak Republic and the Czech Republic, among others (Figure 3.3, Panel B). Yet how much of this is due to

differences in the underlying mental health of each population, and how much is due to cultural differences leading to differential response behaviours for these screening tools?

Some of these cross-country differences could stem from different levels of stigma towards mental health, with countries having lower overall prevalence levels also showing higher levels of stigma (refer to the previous section, and Table 3.2).

Comparisons of the prevalence of mental ill-health can be difficult within countries, as well. Panel A of Figure 3.4 shows that women in 26 European OECD countries are more likely to report higher levels of psychological distress than men, at all stages of their life. Panel B of Figure 3.4 also suggests that white Americans have higher levels of psychological distress than other racial/ethnic groups, and that Asian-Americans have the lowest levels. Research has shown that there are systematic gender differences in self-report bias, as men tend to minimise their symptoms more than women do (Brown et al., 2018[63]). One study also found that men, but not women, reported fewer depressive symptoms when consent forms indicated that a more involved follow-up might occur (Sigmon et al., 2005[75]). A survey on attitudes towards mental health and stigma in Sweden found that women were more likely to report feeling positive attitudes towards those with mental health conditions than did men (Folkhälsomyndigheten, 2022[67]). Therefore how much of the visible difference is due to differences in reporting rather than differences in actual underlying mental health?

Figure 3.4. Are differences in reported outcomes by gender and race/ethnicity due to differences in underlying mental health or to measurement issues?



Note: Scoring information for Panel A: risk for psychological distress is defined as having a score >= 52 on a scale from 0 (least distress) to 100 (most); Panel B: risk for psychological distress is defined as having a score >= 13 on a scale from 0 (least distress) to 24 (most). Refer to Annex 2.B for more information on individual screening tools.

Source: Panel A: OECD calculations based on the 2018 *European Union Statistics on Income and Living Conditions (EU-SILC)* (n.d.[65]), (database), https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions; Panel B : OECD calculations based on University of Michigan (2021[76]), *Panel Study of Income Dynamics* (database), https://psidonline.isr.umich.edu/default.aspx data from 2019 only.

StatLink https://stat.link/3yntk5

To answer these questions on measurement bias and cross-group comparability, researchers assess whether surveys have structural invariance by key socio-demographic characteristics in the process of clinically validating screening tools. Screening tools for symptoms of depression and anxiety, along with

the WHO-5 and SWEMWBS scales for positive mental health, are the tools that have been most frequently validated across numerous settings (e.g. gender, age cohorts and racial/ethnic groups).

In terms of screening tools for specific mental ill-health conditions, both the PHQ-8 and GAD-7 have been found to be free from basic gender and age biases. The PHQ-8 and -9 have been validated across a number of clinical settings, with different age groups, cultures and racial/ethnic groups (El-Den et al., 2018[8]), and the PHQ-2 has been validated for use in youth and adolescents (Richardson et al., 2010[77]). A study of the PHQ-4, limited to the United States, found no structural invariance by gender and age: its findings may extend to other countries with similar population structures, but not necessarily to others with different population structures (Sunderland et al., 2019[33]).

Positive mental health composite scales also perform strongly on age and gender generalisability. The WHO-5 has been shown to have good construct validity for all age groups and has been deemed suitable for children aged 9 and above (Topp et al., 2015[49]). The MHC-SF performs well across sex, age cohorts and education levels (Santini et al., 2020[48]). WEMWBS was originally validated for an adult population but has since been validated for use in youth aged 11 and above (Warwick Medical School, 2021[78]). In the course of validating the 14-item version of WEMWBS, researchers found evidence that two items showed bias for gender; for example, for any level of mental well-being, men were more likely than women to answer positively for the item “I’m feeling more confident” (Stewart-Brown et al., 2009[45]). These two items were removed in the process of creating the 7-item version of the screening tool (SWEMWBS). This short form displays no response rate differences by gender, marital status or household income (Tennant et al., 2007[14]).

Evidence for validity across racial groups is more mixed for all tools, and much of the evidence comes from either the United States or Canada. There is mixed support for cross-cultural invariance of the CES-D’s factor structure across Latino and Anglo-American populations (Crockett et al., 2005[79]; Posner et al., 2001[80]); one study found that item-level modifications were needed for the CES-D when administered to older Hispanic/Latino and Black respondents (El-Den et al., 2018[8]). Other studies also indicate that Asian-American and Armenian-American populations have a different factor structure, higher depressive symptoms, and a tendency to over-endorse positive affect items, in comparison to Anglo-Americans (Iwata and Buka, 2002[81]; Demirchyan, Petrosyan and Thompson, 2011[82]). Research in the United Kingdom implementing the GHQ-12 across diverse racial and ethnic groups found some suggestive evidence of differences by group, requiring further study (Bowe, 2017[83]). A study comparing Korean-American and Anglo-American older adults found that cross-cultural factors may significantly influence the diagnostic accuracy of depression scales and potentially result in the use of different cut-off scores for different populations (Lee et al., 2010[84]). Another study revealed that Black/African-American participants with high GAD symptoms scored lower on the GAD-7 than other participants with similar GAD symptoms (Parkerson et al., 2015[85]; Sunderland et al., 2019[33]).

Measures of self-reported mental health may also be susceptible to bias by racial/ethnic identity. US studies have found that ethnicity appears to moderate the relationship between SRMH and a range of mental health conditions. For example, Black and Hispanic/Latino Americans are more likely to report excellent SRMH than white Americans and show a weaker association between SRMH and service use. A study in Canada found that Asian identity was associated with worse SRMH even after controlling for socio-economic status (Ahmad et al., 2014[51]).

Many screening tools have been translated into multiple languages and used in surveys across the globe. The WHO-5, K10, MHI-5, GAD-7 and WEMWBS have all been translated into a number of languages (Sunderland et al., 2019[33]); the WHO-5, for example, has been translated into more than 30 languages and implemented in surveys in six continents (Topp et al., 2015[49]).[10] WEMWBS has been used across 50 countries and translated into 36 languages (Stewart-Brown, 2021[16]; Warwick Medical School, 2021[78]). Psychometric evaluations for the MHC-SF have also been conducted in many countries (Petrillo et al., 2015[47]; Joshanloo et al., 2013[86]; Guo et al., 2015[46]); however, cross-country comparisons in rates of

flourishing show a high degree of variability, some of which may be driven by measurement issues rather than differences in latent mental health (Santini et al., 2020[48]).

Cultural differences pose significant challenges in establishing uniform definitions and descriptions of mental health and threaten cross-country comparisons (see Box 3.4). Cross-cultural validation refers to whether mental health measures that were originally generated in a given culture are applicable, meaningful and thus equivalent in another culture (Huang and Wong, 2014[87]). Most widely used mental health scales have been developed and validated in high-income, Western and English-speaking populations (e.g. North America, Europe, Australia) and therefore assume a Western understanding of mental disorders and symptoms (Sunderland et al., 2019[33]). This can raise questions as to their applicability to other population groups. For example, a review of 183 published studies on the mental health status of refugees reported that 78% of the findings were based on instruments that were not developed or tested specifically in refugee populations (Hollifield et al., 2002[88]).

Evidence on cross-cultural validation for different tools is mixed. WEMWBS has been validated in 17 different languages and local populations as well as for minority populations within the United Kingdom (Warwick Medical School, 2021[78]). Although the PHQ has been validated in many settings and is considered to be one of the more robust screening tools, one study found that, when applied in middle- or low-income countries, it performed well only in student samples and not in clinical samples, leading researchers to suggest that it should be used only in settings with relatively high rates of literacy (El-Den et al., 2018[8]; Ali, Ryan and De Silva, 2016[32]). Similarly, scoring schemes – i.e. the process of determining what score on a screening tool designates risk for a specific mental issue – are often calibrated based on the US general population, where the initial clinical study took place. The scoring scheme of the Kessler scales was designed to seek out maximum precision in the 90th – 99th percentile of the general population distribution, because of US epidemiological evidence that, in any given year, between 6% and 10% of the US population meet the definition of having a serious mental illness. Therefore, these scoring schemes may not be appropriate for other populations with different structures (Kessler et al., 2002[89]). As another example, the mental health component of the SF-12 is typically scored using US-derived item weights for each response category (Ware et al., 2002[90]). International comparisons have been done in Europe and Australia, which have found these weights to be appropriate (Vilagut et al., 2013[91]), but this does not necessarily extend to other regions.

Research is clearly needed on culturally specific mental health scales developed using a bottom-up and open-ended approach, or with a greater degree of local adaptation, and with further testing of existing scales across different cultures and ethnicities (Sunderland et al., 2019[33]). Furthermore, advances in psychometric models and computational statistics have led to new developments in the administration and scoring of screening tools, which can facilitate cross-cultural analyses.[11] Yet it is important to contextualise the magnitude of these differences. Research using data from the Gallup World Poll covering 150 countries on cross-country differences in measures of positive mental health, including life satisfaction, has found that cultural differences account for only 20% of inter-country variation in outcomes. This 20% includes both the impact of different cultures on outcomes, as well as potential measurement bias, an amount that is small in comparison to the impact of objective conditions – such as income, education and employment (Exton, Smith and Vandendriessche, 2015[92]; OECD, 2013[1]). The impact that these objective life conditions have on mental health is also likely to be larger than that of cultural bias. This does not negate the importance of better designing and validating mental health tools across populations, but it does provide a needed reminder that mental health indicators are informative and useful for policy.

Box 3.7. Key messages: Accuracy across groups

- Differences in attitudes toward mental health can lead to differential reporting across countries, as well as by gender, age and racial/ethnic identity.
- Surveys on stigma and discriminatory views have shown differences in attitudes toward mental health by age and gender.
- In the process of validation, screening tools are tested for biases by age, gender and racial/ethnic group. While most screening tools for specific mental ill-health conditions and most composite scales for positive mental health perform well for age and gender, evidence for race/ethnicity is mixed. More research is needed on the performance of self-reported general mental health questions.
- Survey items must be validated in local populations to ensure their suitability. Validation studies conducted in one geographic area, or in one population group, may not be applicable to other contexts.

How comparable are measures over time?

A key goal of policy makers is to understand trends over time. Is population mental health improving or deteriorating? Do policy interventions lead to visible changes in mental health outcomes? It is therefore important that the accuracy of chosen indicators hold not only cross-sectionally but also over time. There are two complications in measuring mental health over time: (1) behavioural and attitudinal changes towards mental health, leading to different response behaviour; and (2) the fact that many of the screening tools have been validated against clinical diagnoses in cross-sectional studies, which may not provide sufficient evidence that they are sensitive to changes over time.

Attitudes towards mental health have changed over the years, and while stigma and bias remain, progress in reducing them has been made. In recent years, governments across the OECD have pursued public information campaigns, especially centred in schools and educational institutions, to destigmatise mental illness. Even before the COVID-19 pandemic, 12 OECD countries waged national campaigns to improve mental health literacy, and five had regional or local campaigns (OECD, 2021[71]). Initial evidence of the impact is mixed: while some studies show little to no decline in stigma to mental health conditions, especially in the long run (Deady et al., 2020[93]; Walsh and Foster, 2021[94]), others point to an increase in service use, such as visits to psychiatric emergency departments (Cheng et al., 2016[95]). A study in the United Kingdom found that exposure to mental health campaigns may have led to an increase in these symptoms among young people; the research suggests that this was not because of a newfound awareness of pre-existing feelings, but a causal result of increased information about mental illness (Harvey, n.d.[96]). Other early research in this vein posits that awareness campaigns may lead to individuals categorizing their feelings and emotions – which may be mild or moderate – as more concerning indications of mental distress, which may then change their own perceptions and behaviours, thus leading to actual worsening of symptoms (Foulkes and Andrews, 2023[97]).

If anti-stigma campaigns are indeed having their intended impact, then general population attitudes toward poor mental health may be changing, and the average person may feel more comfortable speaking openly and honestly about their mental health. This could distort estimates of mental ill-health prevalence over time. If the general population ten years ago felt less comfortable honestly answering questions on how often they felt “down, depressed or hopeless” over the past two weeks, one might expect higher rates of non-response, or of respondents lying about their true feelings, than today; as a result, one would expect to see the reported prevalence of psychological distress *increase* just because of this change in attitudes.

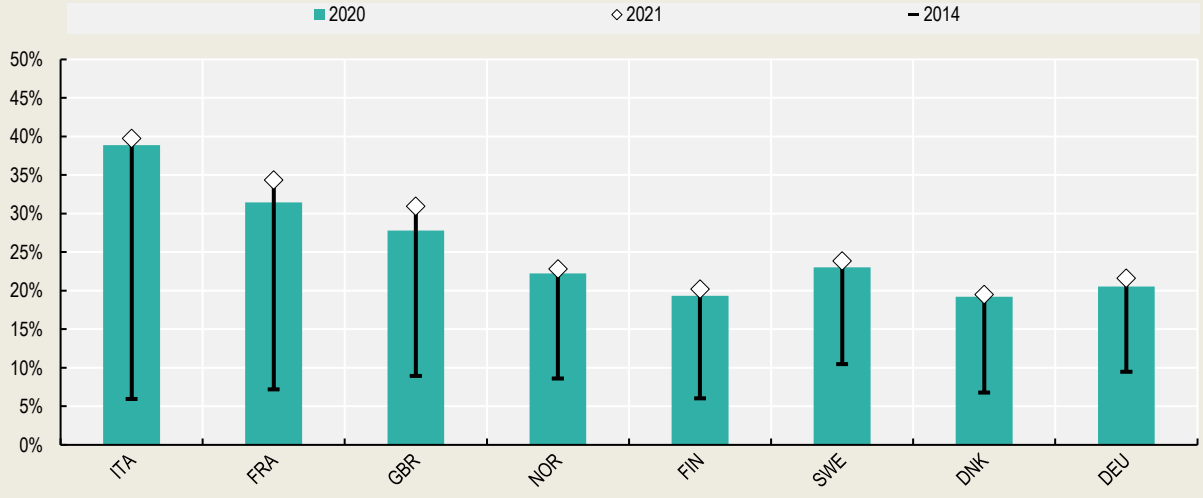
Box 3.8. Changes in mental health during the pandemic

During and in the wake of the COVID-19 pandemic, mental health deteriorated in most OECD countries, with rates of symptoms of depression and anxiety doubling in some (OECD, 2021[98]; OECD, 2021[57]). Indeed, for eight European OECD countries that have comparable pre-pandemic baseline data, the share of the population at risk for depression rose substantially, and by more than 20 percentage points in Italy and France (Figure 3.5; (OECD, 2021[99])). A study looking at data from January 2020 to January 2021 estimated that the share of people experiencing symptoms of anxiety and depression were 28% and 26% higher, respectively, in 2020 than they would have been had the pandemic not occurred (OECD, 2021[57]).[12] Both longitudinal and cross-sectional studies in European countries have found that positive mental health – measured through the WHO-5 (an affect-based measure), and SWEMWBS or the MHC-SF (combining aspects of affect, eudaimonia, social connections and life evaluation) – significantly deteriorated over the course of the pandemic (Thygesen et al., 2021[100]; Vistisen et al., 2022[101]; Eurofound, 2021[102]; Vinko et al., 2022[103]).

While the increase in prevalence of symptoms for mental ill-health is more or less agreed upon, it remains to be seen whether this increase is temporary, or whether mental health will revert to pre-pandemic levels relatively quickly. As of mid-2021, overall mental health had not recovered to pre-pandemic levels; however, there were suggestions of recovery in some OECD member states (OECD, 2021[57]; OECD, 2021[99]). Even still, certain population groups who were particularly negatively affected, such as young people, continue to face many challenges (OECD, 2021[99]).

Figure 3.5. Symptoms of depression rose substantially in eight European OECD countries in the first year of the pandemic

Share of respondents at risk of depression, 2020 and 2021 vs. 2014



Note: Data from 2020 and 2021 come from a different data source than do data from 2014, meaning that caution should be taken in interpreting numerical increases in any individual country. Both data sources use the PHQ-2 as a measure for depression risk. Data for 2020 and 2021 come from the YouGov COVID-19 behaviour tracker: 2020 pooled averages run from April through December, and 2021 pooled averages run from January through June. Baseline data come from the European Health Interview Survey (EHIS) wave 2 in 2014. Refer to Annex 2.B for more information on individual screening tools.
Source: OECD (2021[99]), *COVID-10 and Well-being: Life in the Pandemic*, OECD Publishing, Paris, https://dx.doi.org/10.1787/1e1ecb53-en.

StatLink https://stat.link/szdacx

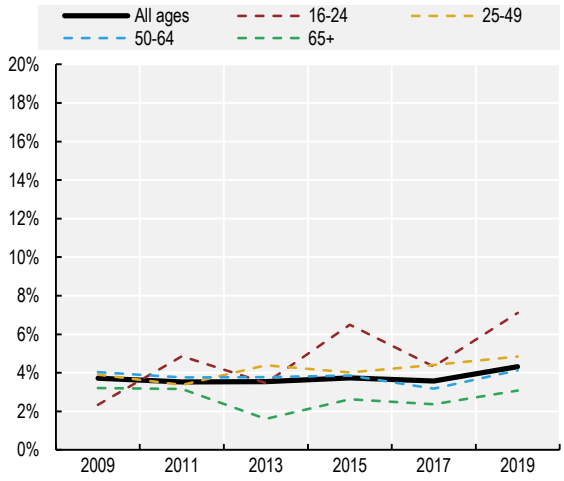
Figure 3.6 provides some evidence disproving this hypothesis. Pre-pandemic data from over 20 European OECD countries and the United States show either improving or stable mental health in the years following the financial crisis and preceding COVID-19 (prevalence of symptoms of anxiety and depression rose dramatically in 2020 at the onset of the pandemic, see Box 3.8). Prevalence of psychological distress in the United States from 2009 to 2019 (measured bi-annually using the K6 screening tool) remained broadly stable over the decade, hovering around 4% (Panel A). Although not controlling for any socio-demographic factors, this suggests that concerns surrounding changing perceptions leading to large changes in response behaviour resulting in higher prevalence rates may not hold. That said, there is some evidence of higher, and potentially rising, prevalence among young people aged 16 to 24, which may reflect a combination of changing circumstances (socio-political, economic, climate-related), changing attitudes among young people toward mental health (making them more willing to speak openly to an enumerator), and smaller sample sizes in this cohort (leading to more noise in the data). Across 26 European OECD countries, psychological distress decreased between 2013 and 2018, which would not be expected if changes in behaviours made respondents more likely to speak honestly about their poor mental health (Panel C); a similar story is shown in Panel D, which shows psychological flourishing in 24 European OECD countries rose between 2011 and 2016. Conversely, data from the United Kingdom (Panel B) show a deterioration of population mental health (as measured by the mental health component of the SF-12), while both their positive mental health (SWEMWBS) and the share at risk for a common mental disorder (GHQ-12) remained more or less stable.

One possible reason why some population surveys may show relatively stable mental health prevalence over time could be that those measures lack sensitivity to change. Many mental health screening tools were validated in cross-sectional clinical samples; researchers therefore caution that they have not been tested for sensitivity to changes over time, which can only be assessed with longitudinal data (Ahmad et al., 2014[51]; Tennant et al., 2007[14]; Moriarty, Zack and Kobau, 2003[21]; Spitzer et al., 2006[11]). However there are exceptions. Some studies have found that the PHQ-8 and PHQ-2 are sensitive to change over time (Löwe et al., 2010[13]). In terms of positive mental health, both WEMWBS and the shorter SWEMWBS have been shown to be sensitive to change over time for both groups and individuals: researchers suggest that +/- 3 points on the WEMWBS scale, and +/- 1 to 3 points on the SWEMWBS scale, indicate a significant change (NHS Health Scotland, 2016[15]; Shah et al., 2018[104]).

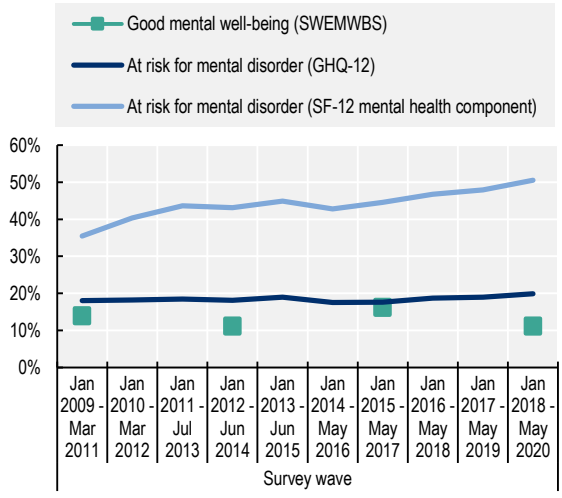
One approach to identifying measurement bias that is driven by changes in individuals' characteristics or circumstances over time is to use longitudinal data. A study by Ploubidis et al. (2019[105]) used two nationally representative surveys in the United Kingdom to track age cohorts over two decades. Using a generalised latent variable measurement modelling framework, researchers tested whether respondents' answers to questions on the Malaise Inventory (a 9-item survey measuring psychological distress) were affected by the passage of time and found little evidence for the presence of bias in the form of age effects, survey design, period effects or cohort-specific effects.

Figure 3.6. Until 2019, mental health improved somewhat in European OECD countries, and remained roughly stable in the United States, despite greater mental health awareness

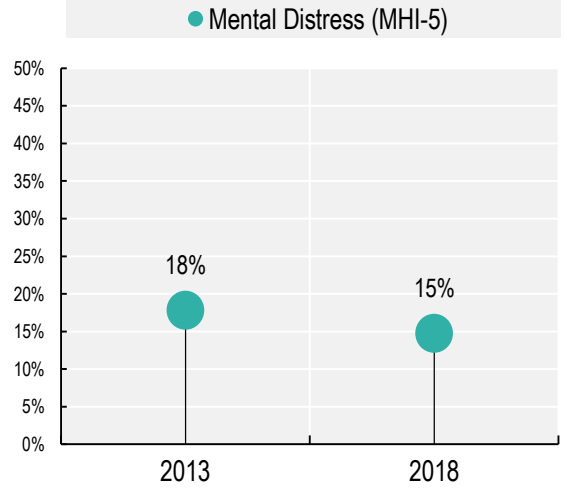
Panel A: Share of the population at risk for mental distress (K6), by age group, USA, 2009-2019



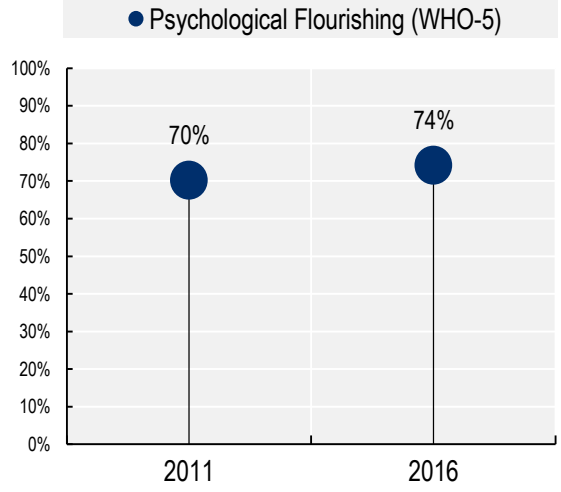
Panel B: Share of the population with good mental health (SWEMWBS) or at risk for a mental disorder (GHQ-12, and mental health component of the SF-12), GBR, 2009-2020



Panel C: Share of the population at risk for mental distress (MHI-5), European OECD 26, 2013 vs. 2018



Panel D: Share of the population who are flourishing (WHO-5), European OECD 24, 2011 vs. 2016



Note: Panel A: Scoring information for each screening tools; risk for psychological distress if score is >= 13 on a scale from 0 (least distress) to 24 (most). Panel B: at risk for a mental condition if score is <= 50 on the transformed SF-12 mental health component composite scale, 0 indicates worst mental health and 100 best possible mental health; risk for a probable common mental disorder (CMD) if score is >= 4 on the GHQ-12, as used in (Woodhead et al., 2012[106]); good mental health is defined as having a SWEMWBS score more than one standard deviation above the sample average. Panel C: risk for psychological distress if score is >= 52 on a scale from 0 (least distress) to 100 (most); Panel D: psychological flourishing if score is >= 14 on a scale from 0 (worst mental health outcome) to 24 (best). Refer to Annex 2.B for more information.
Source: Panel A: OECD calculations based on University of Michigan (2021[76]), *Panel Study of Income Dynamics* (database), https://psidonline.isr.umich.edu/default.aspx; Panel B: OECD calculations based on University of Essex, Institute for Social and Economic Research (2022[70]), *Understanding Society: Waves 1-11, 2009-2020 and Harmonised BHPS: Waves 1-18, 1991-2009* (database). 15th Edition. UK Data Service. SN: 6614, http://doi.org/10.5255/UKDA-SN-6614-16; Panel C: OECD calculations based on the 2013 and 2018 *European Union Statistics on Income and Living Conditions (EU-SILC)* (n.d.[65]), (database), https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions; Panel D: OECD calculations based on the 2011 and 2016 *European Quality of Life Survey* (n.d.[107]), (database), https://www.eurofound.europa.eu/surveys/european-quality-of-life-surveys.

StatLink https://stat.link/y5vgfp

Box 3.9. Key messages: Accuracy over time

- Changes in attitudes towards mental health conditions, and comfort discussing these topics openly, may lead to measurement bias when comparing prevalence rates over time.
- While evidence from impact assessments of anti-stigma campaigns is scarce, pre-pandemic evidence from cross-country trend data does not show the clear increase in reported psychological distress that lower stigma would imply.
- Some mental health screening tools – including the PHQ-8 and (S)WEMWBS – have been found to be sensitive to change over time in longitudinal studies, but other tools have not been subject to sensitivity analysis.
- More research into the presence of bias in the form of age, period or cohort-specific effects for mental health outcomes should be done.

Data collection

The practicalities of data collection can have an important impact on respondent behaviour, affecting the comfort and ease with which they interact with an enumerator and thus answer questions. Whether questions are framed as positive or negative statements, the way in which data are collected (enumerator- vs. self-administered) shapes the quality of the eventual output. Because of the sensitive nature of mental health questions, especially for screening tools that deal with suicide and suicidal ideation, additional protocols should be put in place to ensure both respondent and enumerator safety and well-being.

How does question wording affect respondents' attitudes and response behaviour?

The order in which questions are asked, and the way in which questions are framed, may prime respondents to answer in a certain way. OECD research into subjective well-being shows that the influence of question ordering on life evaluation and affect questions can be significant; because of this, subjective well-being questions should be placed early on in surveys to minimise interference from other modules (OECD, 2013[1]).

For mental health questions, there is some evidence suggesting that questions may be upsetting to respondents, raising ethical concerns. Some studies have shown that participating in a survey with distressing questions, or answering questions focusing on distressing life events, can increase respondents' stress and worsen their mood (Labott et al., 2013[108]), especially among populations already at risk for psychological distress. However, other research into the impacts of mental health surveys on the mood of respondents has not found evidence of significant effects (Jorm et al., 1994[109]; Jacomb et al., 1999[110]). The small portion of interviewees who did report feeling distress were more likely to be young women and people lacking social support (Jacomb et al., 1999[110]).

Within a given mental health screening tool or composite scale, framing questions in a positive or negative light can impact on responses. Some tools use only negative question framing (e.g. PHQ-8, CES-D, K6), some only positive (i.e. (S)WEMWBS, WHO-5), and some employ a mix of the two (e.g. GHQ-12, MHI-5). A negatively framed question might ask, for example, how often someone felt downhearted and depressed, whereas a positively framed question might ask how often someone felt cheery and light-hearted. A respondent may feel more comfortable answering that s/he "rarely" felt cheery, rather than answering that s/he "always" felt depressed. This point is illustrated in Table 3.3, which relies on data from the UKHLS survey. The same sample of respondents were asked questions from three different mental health screening surveys. There are overlaps in the types and content of questions asked; pairs of questions are

showcased in the top portion of the table. The correlation in responses are highest for questions that appear in tools with similar tone framing, either positive or negative (e.g. feeling downhearted and depressed from the mental health component of the SF-12 vs. feeling unhappy and depressed from the GHQ-12), and lowest for items that come from tools that use different framing (e.g. been able to face up to problems from the GHQ-12 and dealing with problems well from SWEMWBS).

Users of mental health services in the United Kingdom have expressed a preference for survey tools that focus on positive, rather than negative, emotions (Stewart-Brown, 2021[16]). A study conducted with users there found that respondents found it "upsetting" to be asked a series of negative items in mental health questionnaires, and they expressed a preference for questions – WEMWEBS, specifically – that focus on aspects of good mental health (Crawford et al., 2011[111]).

Table 3.3. The correlation of answers to similar questions depends on whether the phrasing is positive or negative

Correlation between similarly worded questions on different mental health screening tools

Question phrasing			
	GHQ-12	SF-12	SWEMWBS
	Been able to face up to your problems		Been dealing with problems well
	Been feeling unhappy and depressed	Felt downhearted and depressed	
		Felt calm and peaceful	Been feeling relaxed
Answer correlations			
	Feeling unhappy/depressed (GHQ-12 and SF-12): 0.67	Facing up to problems (GHQ-12 and SWEMWBS): 0.36	Feeling relaxed (SF-12 and SWEMWBS): 0.56

Note: Correlations show the pairwise Pearson correlation coefficient between similarly phrased questions appearing on different mental health screening tools or scales from the same longitudinal survey. Refer to Annex 2.B for more information about specific tools.
Source: OECD calculations based on University of Essex, Institute for Social and Economic Research (2022[70]), *Understanding Society: Waves 1-11, 2009-2020 and Harmonised BHPS: Waves 1-18, 1991-2009* (database). 15th Edition. UK Data Service. SN: 6614, http://doi.org/10.5255/UKDA-SN-6614-16, from wave 10 only (Jan 2018 – May 2020).

Box 3.10. Key messages: Question framing

- OECD research has shown that the ordering of subjective questions in household surveys can influence responses, therefore consistency in ordering across surveys is important, and whenever possible these questions should appear early in surveys.
- Evidence from the United Kingdom's *Understanding Society* survey shows that whether a concept is framed in a negative or a positive light can lead to different responses.
- Some users of mental health services have expressed a preference for survey tools that focus on positive rather than negative emotions.

Does the survey mode affect respondents' answers?

Survey modes, i.e. the way in which data are collected from respondents, can influence how respondents process and reply to questions, as well as how much information they feel comfortable revealing. One of the main drivers of differential responses based on survey mode is social desirability bias: the tendency to

present oneself in a favourable light and/or provide responses that conform to prevailing social norms. Social desirability has two components: impression management and self-deception (Paulhus, 1984[112]). Research has shown that interview subjects under-report taboo topics and over-report socially desirable actions (Krumpal, 2013[59]; Presser and Stinson, 1998[113]). Social desirability bias can present itself in different ways, depending on the way in which data are collected – by an interviewer or self-administered, in person, or over the phone or Internet.

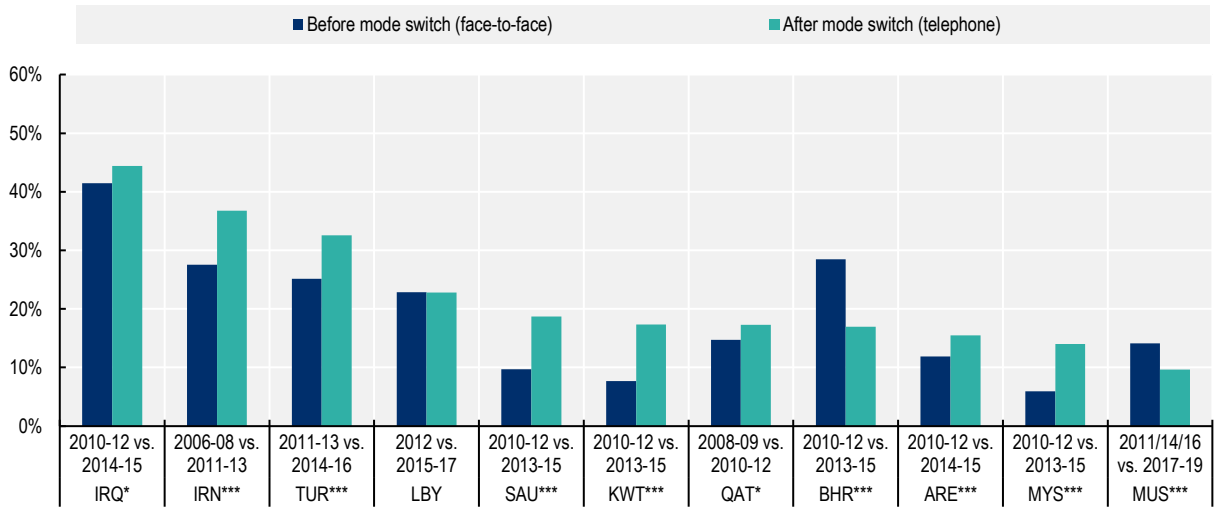
Research has found that respondents are more likely to report better physical and mental health outcomes in interviewer-administered surveys as compared to self-administered surveys. Research has also shown that self-administered survey respondents are less likely to report stigmatised medical conditions, including anxiety and mood disorders (Krumpal, 2013[59]; Latkin et al., 2017[114]). A study in Norway found that respondents were more likely to report symptoms of anxiety and depression in self-administered surveys as compared to interviewer-administered (either in person or over the phone) surveys (Moum, 1998[115]): the presence of social desirability bias was particularly strong for young, well-educated respondents. Another study comparing computer-assisted self-interviewing (ACASI) with interviewer-administered paper-and-pencil (I-PAPI) surveys concluded that respondents were more likely to report mental health symptoms – as measured by the WHO-CIDI – in the self-administered survey than in the interviewer-administered one (Epstein, Barker and Kroutil, 2001[116]), with large differences for major depressive episodes and generalised anxiety disorder.[13]

When surveys are administered by an interviewer, evidence is inconsistent as to whether respondents report better mental health outcomes face-to-face than over the phone or Internet. Evidence from the Canadian Community Health Survey (CCHS) suggests that while some physical health indicators are subject to mode effects, mental health outcomes are not[14] (St-Pierre and Béland, 2004[117]). A study in the United States found the opposite impact, with respondents exhibiting stronger social desirability behaviour in telephone interviews than in person (Holbrook, Green and Krosnick, 2003[118]). Finally, in comparing proctored web-based assessments to paper-and-pencil administration modes, a recent meta-analytic review reported that web-based surveys do not offer an advantage regarding socially desirability in self-report questionnaires, and that the mode of administration does not affect reporting of mental health symptoms (Gnambs and Kaspar, 2017[119]).

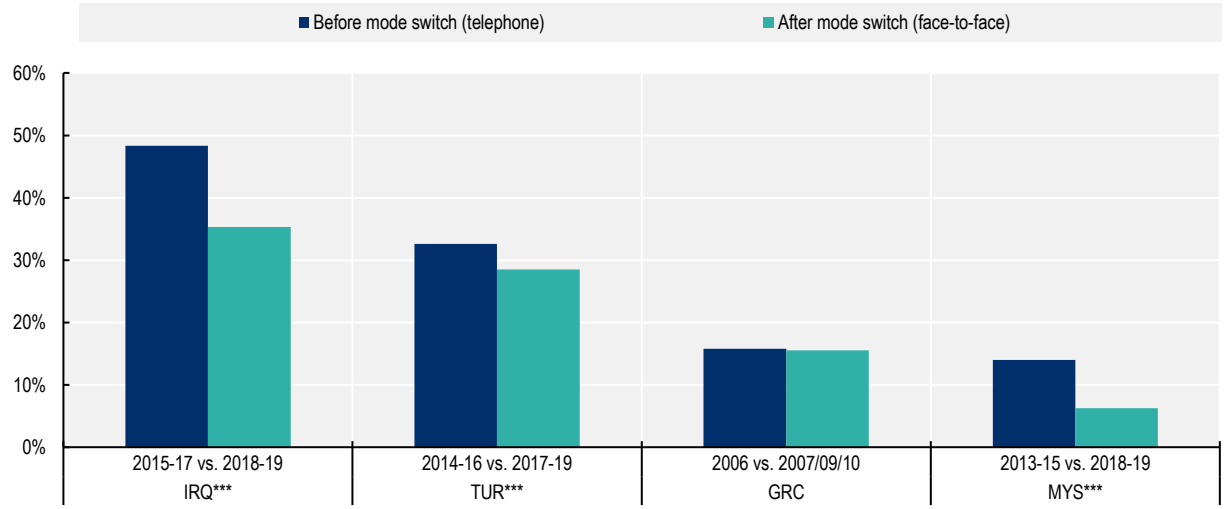
Mode effects in interviewer-administered surveys are illustrated by Figure 3.7. The figure shows the share of the population with a negative affect balance – defined as reporting to have experienced more negative, rather than positive, emotions on the previous day – from Gallup World Poll data. Gallup conducts annual surveys in over 150 countries, including all 38 OECD countries. Data are collected via telephone surveys in many OECD countries; however, face-to-face interviews are common in many places in Latin America, the Middle East, Asia, Africa and former Soviet countries.[15] In a handful of countries, the survey mode changed over the past decade, switching from face-to-face to telephone survey administration and vice versa (indeed, some countries – such as Iraq, Türkiye and Malaysia – have switched multiple times). Figure 3.7 shows that negative affect balance rose (meaning worsening mental health) in eight of 11 countries after switching from in-person to telephone interviews (Panel A); similarly, negative affect balance improved in all four countries after switching from telephone to in-person (Panel B). This is in line with the findings of (Holbrook, Green and Krosnick, 2003[118]), that respondents may be more influenced by social desirability bias when speaking to an interviewer over the phone, and thus over-report good well-being in telephone surveys.[16]

Figure 3.7. Shifts from in-person to telephone-administered surveys are associated with deteriorations in negative affect balance

Panel A: Share of the population with negative affect balance, three-year pooled average before and after mode switch, switch from face-to-face to telephone interviews, 11 countries, varying years



Panel B: Share of the population with negative affect balance, three-year pooled average before and after mode switch, switch from telephone to face-to-face interviews, 4 countries, varying years



Note: Countries followed by *** experienced statistically significant (at the 5% level) changes in outcomes following mode changes; * indicates the change is statistically significant at the 10% level. Three-year averages are shown (three years preceding vs. three years following a mode change); exceptions are made for countries that do not have sufficient years of data collection on either side of a mode switch. In those instances, one- or two-year averages are shown instead. IRQ and ARE did not collect negative affect data in 2013, the year of mode change.
Source: OECD calculations based on the *Gallup World Poll* (n.d.[120]) (database), https://www.gallup.com/178667/gallup-world-poll-work.aspx.

StatLink https://stat.link/ngkd9r

Box 3.11. Key messages: Mode effects

- Respondents are more likely to report worse mental health outcomes when surveys are self-administered, as compared to interviewer-administered.
- When surveys are interviewer-administered, there is conflicting evidence as to whether mental health outcomes are subject to mode effects.
- Consistency in mode is encouraged; when the survey mode changes between data collection rounds, this should be explicitly stated.

What additional protocols or procedures should data collectors take on board for mental health modules?

Interviewer training is crucial to the quality of responses in any survey. However, the measurement of mental health raises additional issues, because of the sensitive nature of the subject matter. Although a body of trained interviewers will generally contribute to higher response rates and better responses, interviewers may struggle to garner responses to questions if they cannot explain adequately to respondents why collecting such information is important and how it will be used. In some cases, respondents may fail to understand why a public agency might want to collect this type of information. To manage risks around respondent attitudes to questions on mental health, it is imperative that interviewers are well-briefed, not just on what concepts the questions are trying to measure, but also on how the information collected will be used. This is essential in order for interviewers to build a strong rapport with respondents, which can help to improve response rates along with the quality of those responses.

A respondent's relationship with the interviewer matters. In a study conducted with a group of mental health service users in a clinical setting, respondents emphasised that a questionnaire was "only as good as the doctor who uses it" (Crawford et al., 2011[111]). In fact, users stated that the interviewer mattered most – more than either the content or length of the survey. A study in the United States found that respondents were more likely to disclose sensitive information about illegal drug use in a face-to-face interview, as opposed to over the phone, with the difference more pronounced for Black compared to white Americans (Aquilino, 1994[121]). A Norwegian study found minimal impact of interviewer gender and age on reported mental health symptoms, but noted that young male interviewers received fewer reports of symptoms as compared to interviewers of other ages and/or female interviewers (Moum, 1998[115]). This can function in the opposite direction as well, with a strong interviewer-respondent bond leading to more information being disclosed.[17]

Recent research has shed more light on the need to involve those with lived experience in the survey design and data collection process, by building a pipeline of researchers with psychiatric disabilities and/or lived experience of mental health conditions (Jones et al., 2021[122]; Banfield et al., 2018[123]; Hancock et al., 2012[124]). There is a strong basis of evidence showing that peer-interviewing techniques – drawing enumerators from the same community as interviewees – can be an effective way of improving trust between interviewer and interviewee, helping to collect high-quality data for hard-to-reach population groups (Dewa et al., 2021[125]; Warr, Mann and Tacticos, 2011[126]; Hancock et al., 2012[124]). Furthermore, in the mental health context, the involvement in research of those with lived mental health experience can improve both the credibility of findings and the likelihood of their adoption into policy (Scholz et al., 2021[127]).

Questions on suicide or suicidal ideation require careful consideration and well-designed procedures to provide needed support to respondents who are at risk (Lakeman and FitzGerald, 2009[128]). The final item of the PHQ-9 asks about suicidal thoughts and ideation, and for this reason it is often excluded from

population surveys. In both the European Health Interview Survey (EHIS) and the United States' National Health Interview Survey (NHIS), the PHQ-8 is used instead, for precisely this reason.

For countries that do include questions on the topic of suicide, additional protocols are often employed. For example, the Australian non-suicidal self-injury (NSSI) prevalence study dealt with the sensitive nature of the survey questions by sharing in advance a large amount of information with the households to be interviewed; this helped to alleviate ethical considerations, and did lead to lower non-response rates (Taylor et al., 2011[129]). In implementing the Mental Health and Access to Care Survey (MHACS), the Canadian government provides mental health resources to both respondents *and* interviewers; enumerators are also provided with employee support services to help them navigate stress or ill effects to their own mental health that may be induced by administering the questionnaire (response to an OECD questionnaire, 2022).

Box 3.12. Key messages: Interviewer training

- Respondents are more likely to participate in an interview, and answer truthfully, if they feel comfortable with the interviewer. Enumerator training should focus on building rapport and trust with respondents.
- Careful procedures and support practices must be in place if surveys are to include questions surrounding suicide and suicidal ideation; in the case of household surveys, it may be best practice to avoid these types of questions on ethical grounds.

Analysis

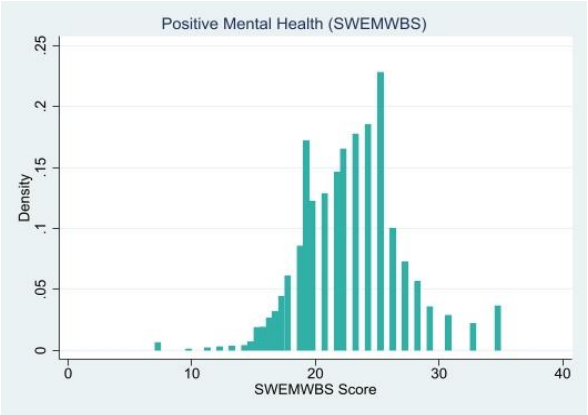
Once mental health data are collected, as accurately and consistently as possible, they are only useful for policy makers when used in analysis. Many of the data described in this report are not binary, meaning the outcome variable is measured on a scale. This is true for the screening tools and composite scales, which contain a number of items, coded and scored accordingly, ranging from worst to best possible mental health. General mental health status tools are typically single questions; however, answer options are not binary and typically use a Likert scale (the exact number and phrasing of answer options varies across scales, see Table 2.11).

What are the trade-offs between using cut-off scores vs. continuous measures?

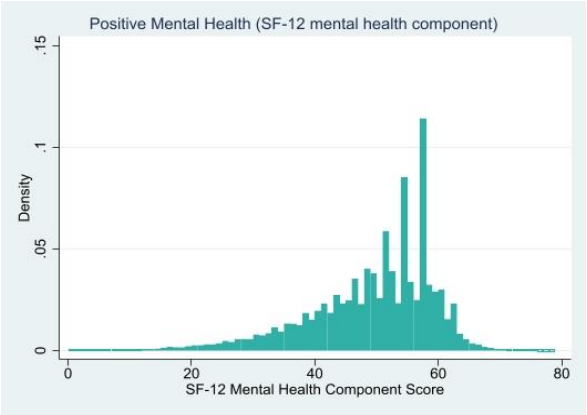
Having a continuous outcome measure for mental health has many benefits, including that it provides more detailed and nuanced information about population mental health, Further, when the distribution of responses is normal, without floor or ceiling effects,[18] this allows for parametric analysis of outcomes, which enables researchers to better analyse the impacts of given policies or interventions. For example, research into screening tools for positive mental health has shown that data sourced from (S)WEMWBS have a distribution that more closely approximates a normal distribution than do screening tools that focus on specific mental illnesses (Shah et al., 2021[17]).[19] Indeed, this can be seen visually in Figure 3.8, which shows density plots for the three mental health tools included in the tenth wave of the UKHLS survey: SWEMWBS; the mental health component of the SF-12 (MHC-12); and the GHQ-12. Of the three, SWEMWBS most closely approximates a normal distribution, followed by the SF-12, while the GHQ-12 shows significant floor effects. Separate research into the MHI-5 has shown that it is positively skewed; it is better able to distinguish between those with worse mental health than between those with higher levels of positive mental health (Elovanio et al., 2020[6]; Thorsen et al., 2013[130]).

Figure 3.8. Positive mental health scales may better approximate a normal distribution

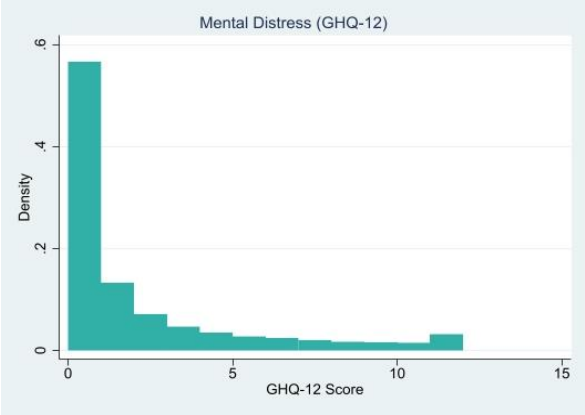
Panel A: Weighted density plot for positive mental health (SWEMWBS), GBR, 2012-2020



Panel B: Weighted density plot for positive mental health (SF-12 mental health component), GBR, 2010-2020



Panel C: Weighted density plot for mental distress (GHQ-12), GBR, 2010-2020



Note: Density plots showing the weighted scores for: Panel A: SWEMWBS (ranging from 9.5 as worse mental well-being, and 35 as better mental well-being); Panel B: the mental health components of the SF-12 (ranging from 0, low functioning, to 100, high functioning); and Panel C: the GHQ-12 (ranging from 0, better mental health, to 12, worse mental health) is lowest. Data for SF-12 and the GHQ-12 are from waves 2 to 10 of *Understanding* Society; data for SWEMWBS come from waves 4, 7 and 10. Refer to Annex 2.B for more information on individual screening tools.

Source: OECD calculations based on University of Essex, Institute for Social and Economic Research (2022[70]), *Understanding Society: Waves 1-11, 2009-2020 and Harmonised BHPS: Waves 1-18, 1991-2009* (database). 15th Edition. UK Data Service. SN: 6614, http://doi.org/10.5255/UKDA-SN-6614-16, from wave 10 only (Jan 2018 – May 2020).

While normal distributions are useful for regression analysis, in order for mental health information to be useful at either a micro-level (i.e. primary care physician, conducting a screening interview to see if a patient is at risk and requires more support) or macro-level (i.e. a government office tasked with tracking changes in risk over time), it is often useful to use cut-off scores to group outcomes into categories. These categories vary depending on the screening tool and scoring convention used, but typically encompass things such as “at risk for depression”, “at risk for anxiety”, “major depressive disorder”, “severe psychological distress”, “psychological flourishing”, etc. These categories can also be useful in analysis to understand how mental health interacts with other aspects of well-being: for example, the share of the employed or unemployed who are experiencing anxiety, or the quality of social connections for those at risk for depression compared to those who are not (OECD, 2021[57]).

One general criticism of the use of thresholds is that they can be arbitrary. However, in the case of mental health screeners, thresholds are established through a rigorous validation process; researchers use receiver operating characteristic (ROC) analysis to determine which cut-off score maximises both the sensitivity and the specificity of the measure (see Box 3.3).[20] Cut-off scores are also useful in that they convert responses to a series of screening tool questions into something comparable to the results of an in-depth diagnostic interview: risk for a certain mental health condition. The PHQ, GAD, Kessler and CES-D surveys all have standard, validated cut-off scores (Kessler et al., 2002[89]; Kroenke et al., 2007[44]; Moriarty, Zack and Kobau, 2003[21]; Manea, Gilbody and McMillan, 2015[131]; Kroenke et al., 2009[12]; Spitzer et al., 2006[11]). The traditional CES-D cut-off score indicative of "depressive case" in clinical samples is 16, but this threshold has been known to produce a high rate of false positives in non-clinical samples (Eaton, 2004[132]; Santor and Coyne, 1997[133]). The GAD-7 also has established cut-off scores, but studies have found that it performs better at identifying the share of the population at risk for generalised anxiety disorder and less well at picking up on other types of anxiety disorders, such as social anxiety disorder (Beard and Björgvinsson, 2014[134]; Sunderland et al., 2019[33]).

Though the GHQ-12 is commonly used to screen general mental health conditions, it has been found to generate a high level of false positives; one study found that as many as half of those identified as having a mental disorder were false positives (positive predictive value of 0.53) (Schmitz, Kruse and Tress, 2001[4]). Other mental ill-health screening tools like the MHI-5 or GHQ-12 were not developed with a standard validated cut-off to define a case of common mental disorder. Although these scales may not have an internationally comparable cut-off score, they have been validated in several studies. For instance, (Berwick et al., 1991[38]) validated the MHI-5 as a measure for depression using clinical interviews as the gold standard and reported an optimal cut-off score of 52.[21] Subsequent research has corroborated the finding that the MHI-5 performs well as a screener for depression and general mood disorders but much less well as a measure for anxiety, somatoform disorders and substance use disorders (Rumpf et al., 2001[135]; Strand et al., 2003[7]; Thorsen et al., 2013[130]).

Some tools have multiple accepted cut-off scores, depending on the intended diagnosis, meaning varying scoring conventions can lead to different prevalence estimates. Figure 3.9 shows the density plot for PHQ-8 scores, ranging from 0 (least at risk for depression) to 25 (most at risk) for 22 European OECD countries. The vertical lines show different validated thresholds. A score of 10 or above indicates risk for major depressive disorder (shown in black) (Kroenke et al., 2008[136]). Other threshold categorisations deem a score of 5-9 as risk for mild depression, 10-14 as moderate, 15-19 as risk for depression, and 20+ as risk for severe depression (Kroenke, Spitzer and Williams, 2001[137]). Another scoring convention (not shown in Figure 3.9), used by Eurostat, is not based purely on the raw score but rather defines major depressive symptoms by respondent answers to individual questions.[22] All three measures lead to different prevalence estimates from the same underlying dataset: (1) 6.9% at risk for major depressive disorder; (2) 15.2% at risk for mild depression, 2.9% at risk for moderate, 1.7% at risk for moderately severe and 0.8% at risk for severe; and (3) 3.1% with major depressive symptoms.[23]

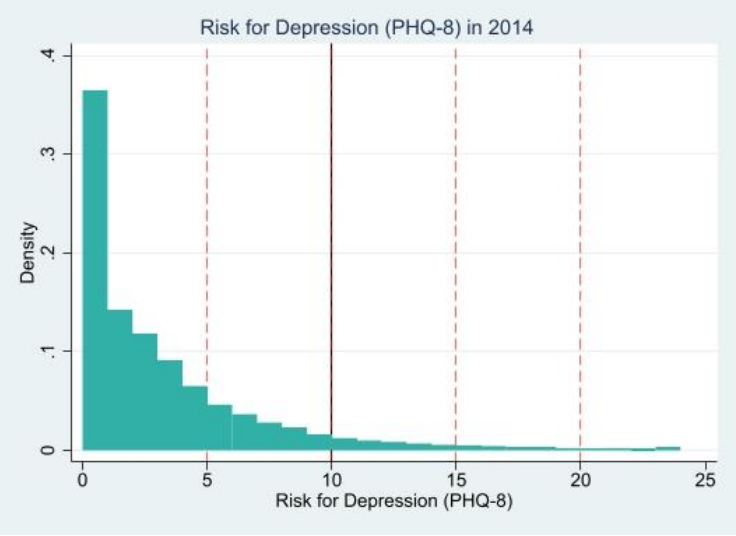
Although there is no clinical gold standard for psychological well-being, positive mental health composite scales have also developed cut-off scores at the request of users. Two main approaches have been put forward for (S)WEMWBS: one statistical and the other benchmarking.[24] In the first, researchers recommend cut-off points at +/- one standard deviation, which result in approximately 15% of the population having high well-being and 15% having low well-being. In the second approach, cut-off scores for (S)WEMWBS are benchmarked against measures capturing symptoms of depression and anxiety (see below for a more detailed discussion of positive mental health tools being used as screeners for mood disorders). Studies have benchmarked WEMWBS against the CES-D, PHQ-9 and GAD-7 to suggest cut-off points on the WEMWBS scale that indicate risk for probable clinical depression, possible depression or mild depression (and/or anxiety). Taking all of this together, researchers suggested that a cut-off point of 60 on the WEMWBS scale, and of 28 on the SWEMWBS scale, can be used to identify the top 15% of

those with high mental well-being, but caution that because there is no clinical measure of high mental well-being these thresholds are by definition arbitrary (Warwick Medical School, 2021[78]).

The MHC-SF comprises three subscales (emotional, social and psychological well-being), which can be scored to group individuals into three categories: flourishing, languishing, and for those in neither of the previous two categories, moderately mentally healthy (Lamers et al., 2011[20]). However surveys in Canada and Denmark have found that the majority of the population scores highly enough to be categorized as flourishing, which runs counter to the theory that flourishing and languishing represent a minority of the population and are deviations from the average. This suggests that more conservative scoring criteria could be warranted to improve the sensitivity of the measure (Santini et al., 2020[48]).

Figure 3.9. Different scoring conventions can lead to different estimates of prevalence

Density plots showing distribution of risk for depression (PHQ-8), European OECD 22, 2014



Note: Weighted density plot for PHQ-8 scores in 22 European OECD countries; scores range from 0 (lowest risk) to 25 (highest risk for depression). Vertical lines indicate validated cut-off scores as established in the literature: As shown by the bold black vertical line, a score >= 10 or above indicates risk for major depressive disorder (Kroenke et al., 2008[136]); as shown by the dotted red line vertical lines, a score of 5-9 as risk for mild depression, 10-14 as moderate, 15-19 as risk for depression, and 20+ as risk for severe depression (Kroenke, Spitzer and Williams, 2001[137]).
Source: OECD calculations based on *European Health Interview Survey* (EHIS) wave 2 data (n.d.[66]) (database), https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:European_health_interview_survey_(EHIS).

Box 3.13. Key messages: Continuous measures vs. cut-off scores

- Mental health tools that provide a continuous, as opposed to binary, outcome variable provide a more nuanced view of population mental health.
- Evidence suggests that positive mental health composite scales better approximate a normal distribution than do measures of psychological distress.
- Cut-off scores provide researchers and policy makers with clear categories of who is at risk and who is not. While cut-off scores for mental ill-health screening tools have been validated against clinical diagnoses to maximise their sensitivity and specificity, no such gold standard exists for psychological flourishing.

- Despite best efforts to ensure sensitivity and specificity, cut-off scores may provide false positives or be ill-suited for some population groups.
- Different scoring conventions for the same screening tool can lead to different prevalence estimates, therefore care should be taken to ensure consistency.

Conclusion

The questions addressed in this chapter are important to consider when thinking about which tools are best for measuring population mental health. As with any survey data, a number of challenges exist, and care is needed when interpreting changes over time and across groups. Perhaps more unique to mental health, stigma and discriminatory views can contribute to bias in reported data. Furthermore, it is important to integrate the perspective of those with lived experience in survey design to ensure the quality and policy relevance of data. However, the evidence reviewed in this chapter shows that existing mental health tools provide useful and policy-relevant outcomes. Given the increasing urgency of the mental health crisis and the prioritisation of action on this front by governments, collecting high-quality mental health data following existing good practice is all the more important. On-going research into open questions of measurement can then progress in tandem with the continual monitoring of population mental health.

All OECD countries are currently measuring population mental health in one form or another and are already making cross-group and longitudinal comparisons. While additional research is needed to test the sensitivity of some tools to change over time, the high-frequency data collected during the COVID-19 pandemic showed that many mental health measures are indeed sensitive to change. Whereas policy discussions prior to COVID-19 sometimes emphasised that rates of common mental health conditions like generalised anxiety disorder and depressive disorders had remained stable in recent years, there is now broad consensus that the pandemic caused a dramatic increase in rates of psychological distress over the first two years – and these spikes have been captured in the data collected in OECD countries.

The task ahead is to better harmonise data collection and provide recommendations for quality improvement for initiatives already underway. The results of a 2022 OECD questionnaire, showcased in Chapter 2, illustrate remaining gaps in the type of mental health outcomes collected by countries: an absence of a harmonised approach to measure symptoms of anxiety, a lack of standardisation in affective and eudaimonic tools, and very uneven use of tools that measure non-depression, non-anxiety types of specific mental health conditions. The recommendations for these areas made below take into account the practical considerations of data collectors, noting the need to keep any new survey items short.

Based on a comparative assessment of the statistical quality of different tools, their response burden and cost (proxied by item length) as well as information on existing data collection practices (Table 3.1), **the report recommends the inclusion of specific mental health outcome measurement tools for national statistical offices to adopt in household, social and health surveys.** These recommendations do not imply the phasing out of other tools that OECD countries are already using to capture population mental health outcomes, particularly with regard to previous diagnoses and experienced symptoms, or measures from the 2013 *OECD Guidelines on Measuring Subjective Well-being* (such as life satisfaction). Rather, they offer a small set of instruments on which a more internationally harmonised set of population mental health outcome indicators could be built:

- *Mental ill-health –priority recommendation:* The Patient Health Questionnaire-4 (**PHQ-4)** could be included in more frequent surveys, alongside the regular collection of the PHQ-8 or PHQ-9 in health surveys. The PHQ-4 measure combines two depression questions from the longer PHQ-9 scale and two anxiety questions from the GAD-7 screening tool. It covers both depression and anxiety, rather than focusing on only one of these two most common mental health conditions. Furthermore, it does so with only four questions, keeping the module relatively short and with a low response

burden. 81% of OECD countries are already implementing the PHQ-8 or PHQ-9, meaning there is trend data to which the depression questions in the PHQ-4 could be linked.[25] The PHQ-8/9 and the GAD-7 could be retained in specific health surveys, while the PHQ-4 could be introduced in general, more frequent surveys, given its shorter length.

- *Positive mental health –recommendation:* Either the **WHO-5** or **SWEMWBS** could be used to measure affective and eudaimonic aspects of positive mental health in a standardised way across countries. These suggestion are mainly based on trends in country measurement practice. The **WHO-5** is a tool for measuring positive affect in that it is relatively short and easy to implement, is included in the OECD's Subjective Well-being Guidelines as an experimental affect module (OECD, 2013[1]), has been translated into many languages, and has been found to be reliable and valid. Although currently used by only 16% of OECD countries, it has been recommended for use by other OECD projects, including as a part of a conceptual framework for measuring the non-financial performance of firms (Siegerink, Shinwell and Žarnic, 2022[138]) and in an effort to use patient-reported indicator surveys (PaRIS) to centre health care delivery on the outcomes that matter to patients (de Bienassis et al., 2021[139]). **SWEMWBS** is a more comprehensive tool, in that it covers affective, eudaimonic, and social connections aspects of positive mental health. This makes it slightly longer than the WHO-5, though only by two questions. SWEMWBS – or the longer 14-question WEMWBS – has been adopted by 19% of OECD countries. For countries already active in subjective well-being or positive mental health measurement, some of the indicator items within SWEMWBS may overlap with existing data collection efforts to measure concepts such as life evaluation and the quantity and quality of social connections (see (OECD, 2020[140]) and (OECD, 2013[1]) for existing OECD recommendations and examples). In these instances, the WHO-5 may be more suitable in that it covers only affect. The topic of measuring affect and eudaimonia specifically will continue to be explored in future OECD workstreams on subjective well-being.
- *General mental health status – recommendation:* A single question about a respondent's **general mental health status** could be included in a range of different surveys across a country's entire data infrastructure system. Single general mental health questions have less of an evidence base compared to established screening tools, but the findings that do exist suggest it is a useful and meaningful measure. Many OECD countries already collect data on self-reported physical health, thus in question framing it will be important to distinguish between self-reported *physical* vs. self-reported *mental* health. Some OECD countries currently collect a general self-reported health measure that captures both physical and mental health; we recommend separating these measures out. In order for this to happen in an internationally comparable way, more research and coordination must happen to align existing country efforts. Canada has been an early adopter of this approach, and its framing as a self-reported mental health (SRMH) question-and-answer option has already been adopted by Chile and Germany; furthermore, much of the existing evidence-base on these types of question has been produced in Canada. Other countries interested in adding such an item to surveys may be interested in using this framing as well: "In general, how is your mental health? Excellent / Very good / Good / Fair / Poor."

Currently, very few countries are using tools to collect information on mental health conditions beyond depression and anxiety, such as substance use disorders, PTSD, obsessive compulsive disorder, eating disorders, bipolar disorder, etc. There are exceptions – these outcomes are covered by all countries that use structured interviews (see Table 2.4), and France and Slovenia, among a few others, have implemented detailed survey modules with tools that capture diagnoses and symptoms of these conditions. There is value in measuring these concepts as distinct conditions, rather than as a part of general mental ill-health, therefore the statistical agenda moving forward could focus on developing recommendations in this space.

As a general point to note, it is more informative, and thus a better use of limited resources, to diversify *across* tool types and mental health outcome measures, rather than to implement a variety of iterations of the same type of tool. For example, rather than implementing a range of different screening tools to capture depression/anxiety across a country's survey infrastructure, it would be of greater use to harmonise *within* tool areas. This might mean choosing a single depression/anxiety screening tool, then supplement it with single-item question tools to capture received diagnoses, experience of symptoms and so on.

Above all, this report has highlighted the importance of precision when communicating outcome measures. Each tool measures a specific, slightly different facet of population mental health. Furthermore, individual tools can be scored in a variety of ways, each leading to different estimates for mental health outcomes. This speaks to the need for greater harmonisation, but also of clearer communication in terms of stating what is meant by mental health and how it is measured. This is all the more important given the rise of mental health to the top of national agendas in the years following the pandemic.

References

Ahmad, F. et al. (2014), "Single item measures of self-rated mental health: A scoping review", *BMC Health Services Research*, Vol. 14/398, pp. 1-11, https://doi.org/10.1186/1472-6963-14-398. [51]

Ahn, J., Y. Kim and K. Choi (2019), "The psychometric properties and clinical utility of the Korean version of GAD-7 and GAD-2", *Frontiers in Psychiatry*, Vol. 127/10, https://doi.org/10.3389/fpsyt.2019.00127. [10]

Ali, G., G. Ryan and M. De Silva (2016), "Validated screening tools for common mental disorders in low and middle income countries: A systematic review", *PLoS ONE*, Vol. 11/6, https://doi.org/10.1371/journal.pone.0156939. [32]

Aquilino, W. (1994), "Interview mode effects in surveys of drug and alcohol use: A field experiment", *Public Opinion Quarterly*, Vol. 58/2, pp. 210-240, https://doi.org/10.1086/269419. [121]

Banfield, M. et al. (2018), "Lived experience researchers partnering with consumers and carers to improve mental health research: Reflections from an Australian initiative", *International Journal of Mental Health Nursing*, Vol. 27/4, pp. 1219-1229, https://doi.org/10.1111/INM.12482. [123]

Beard, C. and T. Björgvinsson (2014), "Beyond generalized anxiety disorder: Psychometric properties of the GAD-7 in a heterogeneous psychiatric sample", *Journal of Anxiety Disorders*, Vol. 28/6, pp. 547-552, https://doi.org/10.1016/j.janxdis.2014.06.002. [134]

Berwick, D. et al. (1991), "Performance of a five-item mental health screening test", *Medical Care*, Vol. 29/2, pp. 169-176, https://doi.org/10.1097/00005650-199102000-00008. [38]

Bharadwaj, P., M. Pai and A. Suziedelyte (2017), "Mental health stigma", *Economics Letters*, Vol. 159, pp. 57-60, https://doi.org/10.1016/j.econlet.2017.06.028. [73]

Böhnke, J. and T. Croudace (2016), "Calibrating well-being, quality of life and common mental disorder items: Psychometric epidemiology in public mental health research", *The British Journal of Psychiatry*, Vol. 209/2, pp. 162-168, https://doi.org/10.1192/BJP.BP.115.165530. [50]

Botha, F., P. Butterworth and R. Wilkins (2021), "Taking the pulse of the nation: Validating a single-item measure of mental distress", *Melbourne Institute Working Paper Series*, No. 6/21, Melbourne Institute of Applied Economic and Social Research, The University of Melbourne, https://melbourneinstitute.unimelb.edu.au/publications/working-papers/search/result?paper=3821299. [52]

Bowe, A. (2017), "The cultural fairness of the 12-item general health questionnaire among diverse adolescents", *Psychological Assessment*, Vol. 29/1, pp. 87-97, https://doi.org/10.1037/PAS0000323. [83]

Brown, S. et al. (2018), "Mental health and reporting bias: Analysis of the GHQ-12", *IZA Discussion Paper* no. 11771, https://www.iza.org/publications/dp/11771/mental-health-and-reporting-bias-analysis-of-the-ghq-12. [63]

Cheng, J. et al. (2016), "Impact of a mass media mental health campaign on psychiatric emergency department visits", *Canadian Journal of Public Health*, Vol. 107/3, https://www.jstor.org/stable/90006480. [95]

Coles, M. and S. Coleman (2010), "Barriers to treatment seeking for anxiety disorders: Initial data on the role of mental health literacy", *Depression and Anxiety*, Vol. 27/1, pp. 63-71, https://doi.org/10.1002/DA.20620. [56]

Cornelius, B. et al. (2013), "The performance of the K10, K6 and GHQ-12 to screen for present state DSM-IV disorders among disability claimants", *BMC Public Health*, Vol. 13/128, http://www.biomedcentral.com/1471-2458/13/128. [37]

Crawford, M. et al. (2011), "Selecting outcome measures in mental health: The views of service users", *Journal of Mental Health*, Vol. 20/4, pp. 336-346, https://doi.org/10.3109/09638237.2011.577114. [111]

Crockett, L. et al. (2005), "Measurement equivalence of the center for epidemiological studies depression scale for Latino and Anglo adolescents: A national study", *Journal of Consulting and Clinical Psychology*, Vol. 73/1, pp. 47-58, https://doi.org/10.1037/0022-006X.73.1.47. [79]

Dadfar, M. et al. (2018), "Reliability, validity, and factorial structure of the World Health Organization-5 Well-Being Index (WHO-5) in Iranian psychiatric outpatients", *Trends in Psychiatry and Psychotherapy*, Vol. 40/2, pp. 79-84, https://doi.org/10.1590/2237-6089-2017-0044. [18]

de Bienassis, K. et al. (2021), "Patient-reported indicators in mental health care: Towards international standards among members of the OECD", *International Journal for Quality in Health Care*, Vol. 6/34(Suppl 1), https://doi.org/10.1093/INTQHC/MZAB020. [139]

de Graaf, R. et al. (2000), "Psychiatric and sociodemographic predictors of attrition in a longitudinal study the Netherlands Mental Health Survey and Incidence Study (NEMESIS)", *American Journal of Epidemiology*, Vol. 152/11, pp. 1039-1047, https://doi.org/10.1093/AJE/152.11.1039. [60]

de Graaf, R., M. Have and S. Van Dorsselaer (2010), "The Netherlands Mental Health Survey and Incidence Study-2 (NEMESIS-2): Design and methods", *International Journal of Methods in Psychiatric Research*, Vol. 19/3, pp. 125-141, https://doi.org/10.1002/MPR.317. [69]

Deady, M. et al. (2020), "Suicide awareness campaigns: Are they a valid prevention strategy?", in *What Can Be Done to Decrease suicidal Behaviour in Australia? A Call to Action*, https://www.blackdoginstitute.org.au/wp-content/uploads/2020/09/What-Can-Be-Done-To-Decrease-Suicide_Chapter-3-Awareness-Campaigns.pdf. [93]

Demirchyan, A., V. Petrosyan and M. Thompson (2011), "Psychometric value of the Center for Epidemiologic Studies Depression (CES-D) scale for screening of depressive symptoms in Armenian population", *Journal of Affective Disorders*, Vol. 133/3, pp. 489-498, https://doi.org/10.1016/j.jad.2011.04.042. [82]

Dere, J. et al. (2015), "Cross-cultural examination of measurement invariance of the Beck Depression Inventory-II", *Psychological Assessment*, Vol. 27/1, pp. 68-81, https://doi.org/10.1037/pas0000026. [149]

Dewa, L. et al. (2021), "Reflections, impact and recommendations of a co-produced qualitative study with young people who have experience of mental health difficulties", *Health Expectations*, Vol. 24/S1, pp. 134-146, https://doi.org/10.1111/HEX.13088. [125]

Dhingra, S. et al. (2011), "PHQ-8 days: A measurement option for DSM-5 Major Depressive Disorder (MDD) severity", *Population Health Metrics*, Vol. 9/11, https://doi.org/10.1186/1478-7954-9-11. [42]

Dowling, N. et al. (2016), "Measurement and control of bias in patient reported outcomes using multidimensional item response theory", *BMC Medical Research Methodology*, Vol. 16/63, https://doi.org/10.1186/s12874-016-0161-z. [150]

Drapeau, A., R. Boyer and F. Diallo (2011), "Discrepancies between survey and administrative data on the use of mental health services in the general population: Findings from a study conducted in Québec", *BMC Public Health*, Vol. 11/837, pp. 1-10, https://doi.org/10.1186/1471-2458-11-837. [72]

Dunn, K. et al. (2009), "Quantification and examination of depression-related mental health literacy", *Journal of Evaluation in Clinical Practice*, Vol. 15/4, pp. 650-653, https://doi.org/10.1111/J.1365-2753.2008.01067.X. [55]

Easton, S. et al. (2017), "The Kessler psychological distress scale: Translation and validation of an Arabic version", *Health and Quality of Life Outcomes*, Vol. 15/1, https://doi.org/10.1186/S12955-017-0783-9. [9]

Eaton, W. (2004), "Center for Epidemiologic Studies Depression scale: Review and revision (CESD and CESD-R)", in M. E. Maruish (ed.), *The Use of Psychological Testing for Treatment Planning and Outcomes Assessment: Instruments for Adults*, Lawrence Erlbaum Associates Publishers, https://psycnet.apa.org/record/2004-14941-011. [132]

Eaton, W. et al. (1992), "Psychopathology and attrition in the epidemiologic catchment area surveys", *American Journal of Epidemiology*, Vol. 135/9, pp. 1051-1059, https://doi.org/10.1093/OXFORDJOURNALS.AJE.A116399. [61]

El-Den, S. et al. (2018), "The psychometric properties of depression screening tools in primary healthcare settings: A systematic review", *Journal of Affective Disorders*, Vol. 225/1, pp. 503-522, https://doi.org/10.1016/j.jad.2017.08.060. [8]

Elovanio, M. et al. (2020), "General Health Questionnaire (GHQ-12), Beck Depression Inventory (BDI-6), and Mental Health Index (MHI-5): Psychometric and predictive properties in a Finnish population-based sample", *Psychiatry Research*, Vol. 289, p. 112973, https://doi.org/10.1016/j.psychres.2020.112973. [6]

Epstein, J., P. Barker and L. Kroutil (2001), "Mode effects in self-reported mental health data", *Public Opinion Quarterly*, Vol. 65/4, pp. 529-549, https://doi.org/10.1086/323577. [116]

Eurofound (2021), *Living, Working and COVID-19: Mental health and trust decline across EU as pandemic enters another year*, Publications Office of the European Union, https://www.eurofound.europa.eu/topic/covid-19. [102]

Eurofound (n.d.), *European Quality of Life Surveys (EQLS) (database)*, https://www.eurofound.europa.eu/surveys/european-quality-of-life-surveys (accessed on 10 June 2022). [107]

Eurostat (n.d.), *European health interview survey (EHIS)*, https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:European_health_interview_survey_(EHIS) (accessed on 16 June 2021). [66]

Eurostat (n.d.), *European Health Interview Survey (EHIS): Reference Metadata in Euro SDMX Metadata Structure (ESMS)*, https://ec.europa.eu/eurostat/cache/metadata/en/hlth_det_esms.htm (accessed on 25 March 2022). [74]

Eurostat (n.d.), *European Union Statistics on Income and Living Conditions (EU-SILC) (database)*, https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions (accessed on 10 June 2022). [65]

Exton, C., C. Smith and D. Vandendriessche (2015), "Comparing happiness across the world: Does culture matter?", *OECD Statistics Working Papers*, No. 2015/4, OECD Publishing, Paris, https://doi.org/10.1787/5jrqppzd9bs2-en. [92]

Folkhälsomyndigheten (2022), *Synen på psykisk ohälsa och suicid: En befolkningsundersökning om kunskaper och attityder [The view of mental illness and suicide: A population survey on knowledge and attitudes]*, https://www.folkhalsomyndigheten.se/publikationer-och-material/publikationsarkiv/s/synen-pa-psykisk-ohalsa-och-suicid-/?pub=105538#105580. [67]

Foulkes, L. and J. Andrews (2023), "Are mental health awareness efforts contributing to the rise in reported mental health problems? A call to test the prevalence inflation hypothesis", *New Ideas in Psychology*, Vol. 69, p. 101010, https://doi.org/10.1016/J.NEWIDEAPSYCH.2023.101010. [97]

Fried, E. (2017), "The 52 symptoms of major depression: Lack of content overlap among seven common depression scales", *Journal of Affective Disorders*, Vol. 208, pp. 191-197, https://doi.org/10.1016/j.jad.2016.10.019. [146]

Furukawa, T. and D. Goldberg (1999), "Cultural invariance of likelihood ratios for the General Health Questionnaire", *Lancet*, Vol. 353/9152, pp. 561-562, https://doi.org/10.1016/S0140-6736(98)05470-1. [145]

Furukawa, T. et al. (2001), "Stratum-specific likelihood ratios of two versions of the General Health Questionnaire", *Psychological Medicine*, Vol. 31/3, pp. 519-529, https://doi.org/10.1017/s0033291701003713. [144]

Furukawa, T. et al. (2003), "The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of Mental Health and Well-Being", *Psychological Medicine*, Vol. 33/2, pp. 357-362, https://doi.org/10.1017/S0033291702006700. [36]

Gallup (n.d.), *Gallup World Poll (database)*, https://www.gallup.com/analytics/318875/global-research.aspx (accessed on 18 June 2021). [120]

García, M. et al. (2005), "Comparison between telephone and self-administration of Short Form Health Survey Questionnaire (SF-36).", *Gaceta sanitaria / S.E.S.P.A.S*, Vol. 19/6, pp. 433-439, https://doi.org/10.1016/S0213-9111(05)71393-5. [143]

Garland, A. et al. (2018), "Use of the WHO's perceived Well-Being Index (WHO-5) as an efficient and potentially valid screen for depression in a low income country", *Families, Systems and Health*, Vol. 36/2, pp. 148-158, https://doi.org/10.1037/FSH0000344. [19]

Gibbons, C. and S. Skevington (2018), "Adjusting for cross-cultural differences in computer-adaptive tests of quality of life", *Quality of Life Research*, Vol. 27/4, pp. 1027-1039, https://doi.org/10.1007/s11136-017-1738-7. [141]

Gill, S. et al. (2007), "Validity of the mental health component scale of the 12-item Short-Form Health Survey (MCS-12) as measure of common mental disorders in the general population", *Psychiatry Research*, Vol. 152/1, pp. 63-71, https://doi.org/10.1016/j.psychres.2006.11.005. [26]

Gnambs, T. and K. Kaspar (2017), "Socially desirable responding in web-based questionnaires: A meta-analytic review of the candor hypothesis", *Assessment*, Vol. 24/6, pp. 746-762, https://doi.org/10.1177/1073191115624547. [119]

Goldberg, D., T. Oldehinkel and J. Ormel (1998), "Why GHQ threshold varies from one place to another", *Psychol Med*, Vol. 28/4, pp. 915-921, https://doi.org/10.1017/s0033291798006874. [151]

Guo, C. et al. (2015), "Psychometric evaluation of the Mental Health Continuum-Short Form (MHC-SF) in Chinese adolescents: A methodological study", *Health and Quality of Life Outcomes*, Vol. 13/198, https://doi.org/10.1186/S12955-015-0394-2. [46]

Hancock, N. et al. (2012), "Participation of mental health consumers in research: Training addressed and reliability assessed", *Australian Occupational Therapy Journal*, Vol. 59/3, pp. 218-224, https://doi.org/10.1111/J.1440-1630.2012.01011.X. [124]

Hanmer, J., R. Hays and D. Fryback (2007), "Mode of administration is important in US national estimates of health-related quality of life", *Medical Care*, Vol. 45/12, pp. 1171-1179, https://doi.org/10.1097/MLR.0b013e3181354828. [142]

Hapke, U. et al. (2022), "Depressive symptoms in the general population before and in the first year of the COVID-19 pandemic: Results of the GEDA 2019/2020 study", *Journal of Health Monitoring*, Vol. 7/4, https://doi.org/10.25646/10664. [158]

Haroz, E. et al. (2017), "How is depression experienced around the world? A systematic review of qualitative literature", *Social Science & Medicine*, Vol. 183, pp. 151-162, https://doi.org/10.1016/j.socscimed.2016.12.030. [31]

Harvey, S. (n.d.), *Do mental health awareness campaigns work? Let's look at the evidence*, Black Dog Institute, https://www.blackdoginstitute.org.au/news/do-mental-health-awareness-campaigns-work-lets-look-at-the-evidence/ (accessed on 25 March 2022). [96]

Hinshaw, S. and A. Stier (2008), "Stigma as related to mental disorders", *Annual Review of Clinical Psychology*, Vol. 4, pp. 367-393, https://doi.org/10.1146/annurev.clinpsy.4.022007.141245. [53]

Hoeymans, N. et al. (2004), "Measuring mental health of the Dutch population: A comparison of the GHQ-12 and the MHI-5", *Health and Quality of Life Outcomes*, Vol. 2/23, https://doi.org/10.1186/1477-7525-2-2. [154]

Holbrook, A., M. Green and J. Krosnick (2003), "Telephone versus face-to-face interviewing of national probability samples with long questionnaires: Comparisons of respondent satisficing and social desirability response bias", *Public Opinion Quarterly*, Vol. 67/1, pp. 79-125, https://doi.org/10.1086/346010. [118]

Hollifield, M. et al. (2002), "Measuring trauma and health status in refugees: A critical review", *JAMA*, Vol. 288/5, pp. 611-621, https://doi.org/10.1001/jama.288.5.611. [88]

Huang, F. et al. (2006), "Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients", *Journal of General Internal Medicine*, Vol. 21/6, pp. 547-552, https://doi.org/10.1111/j.1525-1497.2006.00409.x. [28]

Huang, W. and S. Wong (2014), "Cross-Cultural Validation", in *Encyclopedia of Quality of Life and Well-Being Research*, Springer Netherlands, https://doi.org/10.1007/978-94-007-0753-5_630. [87]

Iwata, N. and S. Buka (2002), "Race/ethnicity and depressive symptoms: A cross-cultural/ethnic comparison among university students in East Asia, North and South America", *Social Science and Medicine*, Vol. 55/12, pp. 2243-2252, https://doi.org/10.1016/S0277-9536(02)00003-5. [81]

Jacomb, P. et al. (2002), "Predictors of refusal to participate: A longitudinal health survey of the elderly in Australia", *BMC Public Health*, Vol. 2/4, https://doi.org/10.1186/1471-2458-2-4. [68]

Jacomb, P. et al. (1999), "Emotional response of participants to a mental health survey", *Social Psychiatry and Psychiatric Epidemiology*, Vol. 34/2, pp. 80-84, https://doi.org/10.1007/S001270050115. [110]

Jin, K. and W. Wang (2014), "Generalized IRT models for extreme response style", *Educational and Psychological Measurement*, Vol. 74/1, pp. 116-138, https://doi.org/10.1177/0013164413498876. [152]

Jones, N. et al. (2021), "Lived experience, research leadership, and the transformation of mental health services: Building a researcher pipeline", *Psychiatric Services*, Vol. 72/5, pp. 591-593, https://doi.org/10.1176/appi.ps.202000468. [122]

Jorm, A. et al. (1994), "Do mental health surveys disturb? Further evidence", *Psychological Medicine*, Vol. 24/1, pp. 233-237, https://doi.org/10.1017/S0033291700026994. [109]

Joshanloo, M. et al. (2013), "Measurement invariance of the Mental Health Continuum-Short Form (MHC-SF) across three cultural groups", *Personality and Individual Differences*, Vol. 55/7, pp. 755-759, https://doi.org/10.1016/J.PAID.2013.06.002. [86]

Kessler, R. et al. (2002), "Short screening scales to monitor population prevalences and trends in non-specific psychological distress", *Psychological Medicine*, Vol. 32/6, pp. 959-976, https://doi.org/10.1017/S0033291702006074. [89]

Kessler, R. et al. (2004), "The US National Comorbidity Survey Replication (NCS-R): Design and field procedures", *International Journal of methods in Psychiatric Research*, Vol. 13/2, pp. 69-92, https://doi.org/10.1002/MPR.167. [64]

Knudsen, A. et al. (2021), "Prevalence of mental disorders, suicidal ideation and suicides in the general population before and during the COVID-19 pandemic in Norway: A population-based repeated cross-sectional analysis", *The Lancet Regional Health Europe*, Vol. 4, https://doi.org/10.1016/J.LANEPE.2021.100071. [160]

Kroenke, K., R. Spitzer and J. Williams (2001), "The PHQ-9: Validity of a brief depression severity measure", *Journal of General Internal Medicine*, Vol. 16/9, https://doi.org/10.1046/j.1525-1497.2001.016009606.x. [137]

Kroenke, K. et al. (2009), "An ultra-brief screening scale for anxiety and depression: The PHQ-4", *Psychosomatics*, Vol. 50/6, pp. 613-621, https://doi.org/10.1176/appi.psy.50.6.613. [12]

Kroenke, K. et al. (2007), "Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and detection", *Annals of Internal Medicine*, Vol. 146/5, pp. 317-325, https://doi.org/10.7326/0003-4819-146-5-200703060-00004. [44]

Kroenke, K. et al. (2008), "The PHQ-8 as a measure of current depression in the general population", *Journal of Affective Disorders*, Vol. 114/1-3, pp. 163-173, https://doi.org/10.1016/j.jad.2008.06.026. [136]

Krumpal, I. (2013), "Determinants of social desirability bias in sensitive surveys: A literature review", *Quality and Quantity*, Vol. 47/4, pp. 2025-2047, https://doi.org/10.1007/S11135-011-9640-9. [59]

Labott, S. et al. (2013), "Emotional risks to respondents in survey research: Some empirical evidence", *Journal of Empirical Research on Human Research Ethics*, Vol. 8/4, p. 53, https://doi.org/10.1525/JER.2013.8.4.53. [108]

Lakeman, R. and M. FitzGerald (2009), "The ethics of suicide research", *Crisis*, Vol. 30/1, pp. 13-19, https://doi.org/10.1027/0227-5910.30.1.13. [128]

Lamers, S. et al. (2011), "Evaluating the psychometric properties of the Mental Health Continuum-Short Form (MHC-SF)", *Journal of Clinical Psychology*, Vol. 67/1, pp. 99-110, https://doi.org/10.1002/jclp.20741. [20]

Latkin, C. et al. (2017), "The relationship between social desirability bias and self-reports of health, substance use, and social network factors among urban substance users in Baltimore, Maryland", *Addictive Behaviors*, Vol. 73, pp. 133-136, https://doi.org/10.1016/J.ADDBEH.2017.05.005. [114]

Lee, J. et al. (2010), "Cross-cultural considerations in administering the Center for Epidemiologic Studies Depression Scale", *Gerontology*, Vol. 57/5, pp. 455-461, https://doi.org/10.1159/000318030. [84]

Leong, F., P. Priscilla Lui and Z. Kalibatseva (2019), "Multicultural Issues in Clinical Psychological Assessment", in Selbom, M. and J. Suhr (eds.), *The Cambridge Handbook of Clinical Assessment and Diagnosis*, Cambridge University Press, https://doi.org/10.1017/9781108235433. [23]

Löwe, B. et al. (2010), "A 4-item measure of depression and anxiety: Validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population", *Journal of Affective Disorders*, Vol. 122/1-2, pp. 86-95, https://doi.org/10.1016/j.jad.2009.06.019. [13]

Lowthian, P. and L. Lloyd (2020), *How face-to-face interviewer attitudes and beliefs moderate the effect of monetary incentive on UK Labour Force Survey response rates*, Office for National Statistics (ONS), https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/methodologies/howfacetofaceinterviewerattitudesandbeliefsmoderatetheeffectofmonetaryincentiveonuklabourforcesurveyresponserates. [162]

Manea, L., S. Gilbody and D. McMillan (2015), "A diagnostic meta-analysis of the Patient Health Questionnaire-9 (PHQ-9) algorithm scoring method as a screen for depression", *General Hospital Psychiatry*, Vol. 37/1, pp. 67-75, https://doi.org/10.1016/j.genhosppsych.2014.09.009. [131]

Mauz, E. et al. (2022), "Time trends of mental health indicators in Germany's adult population before and during the COVID-19 pandemic", *medRxiv*, https://doi.org/10.1101/2022.10.09.22280826. [159]

Medina-Mora, M. et al. (2008), "The Mexican National Comorbidity Survey (M-NCS): Overview and results", in Kessler, R. and T. Üstün (eds.), *The WHO World Mental Health Surveys: Global Perspectives on the Epidemiology of Mental Disorders*, Cambridge University Press, https://digitallibrary.un.org/record/700329?ln=en. [153]

Moriarty, D., M. Zack and R. Kobau (2003), "The Centers for Disease Control and Prevention's healthy days measures: Population tracking of perceived physical and mental health over time", *Health and Quality of Life Outcomes*, Vol. 1/37, https://doi.org/10.1186/1477-7525-1-37. [21]

Mostafa, T. et al. (2021), "Missing at random assumption made more plausible: Evidence from the 1958 British birth cohort", *Journal of Clinical Epidemiology*, Vol. 136, pp. 44-54, https://doi.org/10.1016/J.JCLINEPI.2021.02.019. [62]

Moum, T. (1998), "Mode of administration and interviewer effects in self-reported symptoms of anxiety and depression", *Social Indicators Research*, Vol. 45/1-3, pp. 279-318, https://doi.org/10.1023/a:1006958100504. [115]

Murphy, R. and B. Hallahan (2016), "Differences between DSM-IV and DSM-5 as applied to general adult psychiatry", *Irish Journal of Psychological Medicine*, Vol. 33/3, pp. 135-141, https://doi.org/10.1017/IPM.2015.54. [30]

New Zealand Government (2018), *He Ara Oranga : Report of the Government Inquiry into Mental Health and Addiction*, Mental Health and Addiction Inquiry, https://mentalhealth.inquiry.govt.nz/inquiry-report/he-ara-oranga/. [34]

NHS Health Scotland (2016), *Warwick-Edinburgh Mental Well-being Scale (WEMWBS): User guide-Version 2*, NHS Health Scotland, Edinburgh, https://s3.amazonaws.com/helpscout.net/docs/assets/5f97128852faff0016af3a34/attachments/5fe10a9eb624c71b7985b8f3/WEMWBS-Scale.pdf. [15]

NHS Health Scotland (2008), *Review of scales of positive mental health validated for use with adults in the UK: Technical report*, NHS Health Scotland, Edinburgh, http://www.healthscotland.scot/media/2244/review-of-scales-of-positive-mental-health-validated-for-use-with-adults-in-the-uk.pdf. [3]

O'Connor, D. and R. Parslow (2010), "Mental health scales and psychiatric diagnoses: Responses to GHQ-12, K-10 and CIDI across the lifespan", *Journal of Affective Disorders*, Vol. 121/3, pp. 263-267, https://doi.org/10.1016/j.jad.2009.06.038. [27]

OECD (2021), *A New Benchmark for Mental Health Systems: Tackling the Social and Economic Costs of Mental Ill-Health*, OECD Health Policy Studies, OECD Publishing, Paris, https://doi.org/10.1787/4ed890f6-en. [71]

OECD (2021), *COVID-19 and Well-being: Life in the Pandemic*, OECD Publishing, Paris, https://doi.org/10.1787/1e1ecb53-en. [99]

OECD (2021), *Fitter Minds, Fitter Jobs: From Awareness to Change in Integrated Mental Health, Skills and Work Policies*, Mental Health and Work, OECD Publishing, https://doi.org/10.1787/a0815d0f-en. [57]

OECD (2021), "Tackling the mental health impact of the COVID-19 crisis: An integrated, whole-of-society response", *OECD Policy Responses to Coronavirus (COVID-19)*, https://www.oecd.org/coronavirus/policy-responses/tackling-the-mental-health-impact-of-the-covid-19-crisis-an-integrated-whole-of-society-response-0ccafa0b/. [98]

OECD (2020), *How's Life? 2020: Measuring Well-being*, OECD Publishing, Paris, https://doi.org/10.1787/9870c393-en. [140]

OECD (2017), *OECD Guidelines on Measuring Trust*, OECD Publishing, Paris, https://doi.org/10.1787/9789264278219-en. [2]

OECD (2013), *OECD Guidelines on Measuring Subjective Well-being*, OECD Publishing, Paris, https://doi.org/10.1787/9789264191655-en. [1]

Ohno, S. et al. (2017), "Smallest detectable change and test-retest reliability of a self-reported outcome measure: Results of the Center for Epidemiologic Studies Depression Scale, General Self-Efficacy Scale, and 12-item General Health Questionnaire", *Journal of Evaluation in Clinical Practice*, Vol. 23/6, pp. 1348-1354, https://doi.org/10.1111/JEP.12795. [5]

Parkerson, H. et al. (2015), "Cultural-based biases of the GAD-7", *Journal of Anxiety Disorders*, Vol. 31, pp. 38-42, https://doi.org/10.1016/j.janxdis.2015.01.005. [85]

Paulhus, D. (1984), "Two-component models of socially desirable responding", *Journal of Personality and Social Psychology*, Vol. 46/3, pp. 598-609, https://psycnet.apa.org/doi/10.1037/0022-3514.46.3.598. [112]

Pedrelli, P. et al. (2013), "Reliability and validity of the Symptoms of Depression Questionnaire (SDQ)", *CNS Spectrums*, Vol. 19/6, pp. 535-546, https://doi.org/10.1017/S1092852914000406. [147]

Petrillo, G. et al. (2015), "The Mental Health Continuum–Short Form (MHC–SF) as a measure of well-being in the Italian context", *Social Indicators Research*, Vol. 121, pp. 291-312, https://doi.org/10.1007/s11205-014-0629-3. [47]

Pettersson, A. et al. (2015), "Which instruments to support diagnosis of depression have sufficient accuracy? A systematic review", *Nordic Journal of Psychiatry*, Vol. 69/7, pp. 497-508, https://doi.org/10.3109/08039488.2015.1008568. [41]

Pintea, S. and R. Moldovan (2009), "The Receiver-Operating Characteristic (ROC) analysis: Fundamentals and applications in clinical psychology", *Journal of Cognitive and Behavioral Psychotherapies*, Vol. 9/1, pp. 49-66, https://psycnet.apa.org/record/2009-04815-004 (accessed on 0). [24]

Ploubidis, G., E. McElroy and H. Moreira (2019), "A longitudinal examination of the measurement equivalence of mental health assessments in two British birth cohorts", *Longitudinal and Life Course Studies*, Vol. 10/4, pp. 471-489, https://doi.org/10.1332/175795919X15683588979486. [105]

Posner, S. et al. (2001), "Factor variability of the Center for Epidemiological Studies Depression Scale (CES-D) among urban Latinos", *Ethnicity and Health*, Vol. 6/2, pp. 137-144, https://doi.org/10.1080/13557850120068469. [80]

Presser, S. and L. Stinson (1998), "Data collection mode and social desirability bias in self-reported religious attendance", *American Sociological Review*, Vol. 63/1, pp. 137-145, https://doi.org/10.2307/2657486. [113]

Public Health Agency of Sweden (2022), *Ny nationell strategi för psykisk hälsa och suicidprevention*, https://www.folkhalsomyndigheten.se/livsvillkor-levnadsvanor/psykisk-halsa-och-suicidprevention/psykisk-halsa/nationell-strategi/. [35]

Richardson, L. et al. (2010), "Evaluation of the Patient Health Questionnaire-9 item for detecting major depression among adolescents", *Pediatrics*, Vol. 126/6, pp. 1117-1123, https://doi.org/10.1542/peds.2010-0852. [77]

Rivera-Riquelme, M., J. Piqueras and P. Cuijpers (2019), "The revised Mental Health Inventory-5 (MHI-5) as an ultra-brief screening measure of bidimensional mental health in children and adolescents", *Psychiatry Research*, Vol. 274, pp. 247-253, https://doi.org/10.1016/J.PSYCHRES.2019.02.045. [40]

Rose, M. and J. Devine (2014), "Assessment of patient-reported symptoms of anxiety", *Dialogues in Clinical Neuroscience*, Vol. 16/2, pp. 197-211, https://doi.org/10.31887/DCNS.2014.16.2/mrose. [43]

Rumpf, H. et al. (2001), "Screening for mental health: Validity of the MHI-5 using DSM-IV Axis I psychiatric disorders as gold standard", *Psychiatry Research*, Vol. 105/3, pp. 243-253, https://doi.org/10.1016/S0165-1781(01)00329-8. [135]

Santini, Z. et al. (2022), "Higher levels of mental wellbeing predict lower risk of common mental disorders in the Danish general population", *Mental Health & Prevention*, Vol. 26, p. 200233, https://doi.org/10.1016/j.mhp.2022.200233. [148]

Santini, Z. et al. (2020), "Measuring positive mental health and flourishing in Denmark: Validation of the mental health continuum-short form (MHC-SF) and cross-cultural comparison across three countries", *Health and Quality of Life Outcomes*, Vol. 18/297, https://doi.org/10.1186/s12955-020-01546-2. [48]

Santomauro, D. et al. (2021), "Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic", *The Lancet*, Vol. 398/10312, pp. 1700-1712, https://doi.org/10.1016/S0140-6736(21)02143-7. [161]

Santor, D. and J. Coyne (1997), "Shortening the CES-D to improve its ability to detect cases of depression", *Psychological Assessment*, Vol. 9/3, pp. 233-243, https://doi.org/10.1037/1040-3590.9.3.233. [133]

Schmitz, N., J. Kruse and W. Tress (2001), "Improving screening for mental disorders in the primary care setting by combining the GHQ-12 and SCL-90-R subscales", *Comprehensive Psychiatry*, Vol. 42/2, pp. 166-173, https://doi.org/10.1053/COMP.2001.19751. [4]

Scholz, B. et al. (2021), "'People just need to try it to be converted!': A picture of consumer mental health research in Australia and New Zealand", *Issues in Mental Health Nursing*, Vol. 42/3, pp. 249-255, https://doi.org/10.1080/01612840.2020.1795763. [127]

Shah, N. et al. (2021), "Short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS): Performance in a clinical sample in relation to PHQ-9 and GAD-7", *Health and Quality of Life Outcomes*, Vol. 19/260, https://doi.org/10.1186/s12955-021-01882-x. [17]

Shah, N. et al. (2018), "Responsiveness of the Short Warwick Edinburgh Mental Well-Being Scale (SWEMWBS): Evaluation a clinical sample", *Health and Quality of Life Outcomes*, Vol. 16/1, https://doi.org/10.1186/s12955-018-1060-2. [104]

Siegerink, V., M. Shinwell and Ž. Žarnic (2022), "Measuring the non-financial performance of firms through the lens of the OECD Well-being Framework: A common measurement framework for "Scope 1" Social performance", *OECD Papers on Well-being and Inequalities*, No. 3, OECD Publishing, Paris, https://www.oecd-ilibrary.org/social-issues-migration-health/measuring-the-non-financial-performance-of-firms-through-the-lens-of-the-oecd-well-being-framework_28850c7f-en (accessed on 18 May 2022). [138]

Sigmon, S. et al. (2005), "Gender differences in self-reports of depression: The response bias hypothesis revisited", *Sex Roles*, Vol. 53/5-6, pp. 401-411, https://doi.org/10.1007/s11199-005-6762-3. [75]

Singer, E., H. Hippler and N. Schwarz (1992), "Confidentiality assurances in surveys: Reassurance or threat?", *International Journal of Public Opinion Research*, Vol. 4/3, pp. 256-268, https://doi.org/10.1093/IJPOR/4.3.256. [156]

Singer, E., D. Von Thurn and E. Miller (1995), "Confidentiality assurances and response: A quantitative review of the experimental literature", *Public Opinion Quarterly*, Vol. 59/1, pp. 66-77, https://doi.org/10.1086/269458. [58]

Smith, D. and B. Oreskes (2019), "Homeless population's mental illness, substance abuse under-reported", *Los Angeles Times*, https://www.latimes.com/california/story/2019-10-07/homeless-population-mental-illness-disability. [157]

Spitzer, R. et al. (2006), "A brief measure for assessing generalized anxiety disorder: The GAD-7", *Archives of Internal Medicine*, Vol. 166/10, pp. 1092-1097, https://doi.org/10.1001/ARCHINTE.166.10.1092. [11]

Statistics Canada (2021), *Measuring Population Mental Health in Canada*, https://www.slideshare.net/StatsCommunications/oecd-wellbeing-and-mental-health-conference-jennifer-ali-statcan. [29]

Stewart-Brown, S. (2021), *15 years on: Insights and reflections on the Warwick-Edinburgh Mental Wellbeing Scales (WEMWBS)*, What Works Wellbeing, https://whatworkswellbeing.org/resources/insights-and-reflections-on-the-warwick-edinburgh-mental-wellbeing-scales-wemwbs/. [16]

Stewart-Brown, S. et al. (2009), "Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a Rasch analysis using data from the Scottish Health Education Population Survey", *Health and Quality of Life Outcomes*, Vol. 7/1, p. 15, https://doi.org/10.1186/1477-7525-7-15. [45]

St-Pierre, M. and Y. Béland (2004), "Mode effects in the Canadian Community Health Survey: A comparison of CAPI and CATI", in *2004 Proceedings of the American Statistical Association Meeting, Survey Research Methods*, American Statistical Association, Toronto, Canada, https://www.statcan.gc.ca/en/statistical-programs/document/3226_D16_T9_V5. [117]

Strand, B. et al. (2003), "Measuring the mental health status of the Norwegian population: A comparison of the instruments SCL-25, SCL-10, SCL-5 and MHI-5 (SF-36)", *Nordic Journal of Psychiatry*, Vol. 57/2, pp. 113-118, https://doi.org/10.1080/08039480310000932. [7]

Streiner, D. and J. Cairney (2007), "What's under the ROC? An introduction to Receiver Operating Characteristics curves", *The Canadian Journal of Psychiatry*, Vol. 52/2, pp. 121-128, https://doi.org/10.1177/070674370705200210 (accessed on 18 May 2022). [25]

Suhr, D. (2006), *Exploratory or Confirmatory Factor Analysis?*, SAS Users Group International (SUGI), San Francisco, CA, https://support.sas.com/resources/papers/proceedings/proceedings/sugi31/200-31.pdf. [22]

Sunderland, M. et al. (2019), "Self-Report Scales for Common Mental Disorders", in *The Cambridge Handbook of Clinical Assessment and Diagnosis*, Cambridge University Press, https://doi.org/10.1017/9781108235433.019. [33]

Tambling, R., C. D'Aniello and B. Russell (2021), "Mental health literacy: A critical target for narrowing racial disparities in behavioral health", *International Journal of Mental Health and Addiction*, https://doi.org/10.1007/s11469-021-00694-w. [54]

Taylor, A. et al. (2011), "Methodological issues associated with collecting sensitive information over the telephone: Experience from an Australian non-suicidal self-injury (NSSI) prevalence study", *BMC Medical Research Methodology*, Vol. 11, https://doi.org/10.1186/1471-2288-11-20. [129]

Tennant, R. et al. (2007), "The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): Development and UK validation", *Health and Quality of Life Outcomes*, Vol. 5/63, https://doi.org/10.1186/1477-7525-5-63. [14]

Thorsen, S. et al. (2013), "The predictive value of mental health for long-term sickness absence: the Major Depression Inventory (MDI) and the Mental Health Inventory (MHI-5) compared", *BMC Medical Research Methodology*, Vol. 13/115, https://doi.org/10.1186/1471-2288-13-115. [130]

Thygesen, L. et al. (2021), "Decreasing mental well-being during the COVID-19 pandemic: A longitudinal study among Danes before and during the pandemic", *Journal of Psychiatric Research*, Vol. 144, pp. 151-157, https://doi.org/10.1016/J.JPSYCHIRES.2021.09.035. [100]

Topp, C. et al. (2015), "The WHO-5 well-being index: A systematic review of the literature", *Psychotherapy and Psychosomatics*, Vol. 84/3, pp. 167-176, https://doi.org/10.1159/000376585. [49]

University of Essex, I. (2022), *Understanding Society: Waves 1-11, 2009-2020 and Harmonised BHPS: Waves 1-18, 1991-2009 (database)*, 5th Edition. UK Data Service., https://doi.org/10.5255/UKDA-SN-6614-16 (accessed on 10 June 2022). [70]

University of Michigan (2021), *Panel Study of Income Dynamics (database)*, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, https://psidonline.isr.umich.edu/default.aspx (accessed on 10 June 2022). [76]

Vaughn-Coaxum, R., P. Mair and J. Weisz (2016), "Racial/ethnic differences in youth depression indicators", *Clinical Psychological Science*, Vol. 4/2, pp. 239-253, https://doi.org/10.1177/2167702615591768. [155]

Vilagut, G. et al. (2013), "The Mental Component of the Short-Form 12 Health Survey (SF-12) as a measure of depressive disorders in the general population: Results with three alternative scoring methods", *Value in Health*, Vol. 16/4, pp. 564-573, https://doi.org/10.1016/j.jval.2013.01.006. [91]

Vinko, M. et al. (2022), "Positive mental health in Slovenia before and during the COVID-19 pandemic", *Frontiers in Public Health*, Vol. 10, p. 3719, https://doi.org/10.3389/fpubh.2022.963545. [103]

Vistisen, H. et al. (2022), "The less depressive state of Denmark following the second wave of the COVID-19 pandemic", *Acta Neuropsychiatrica*, Vol. 34/3, pp. 163-166, https://doi.org/10.1017/NEU.2022.1. [101]

Walsh, D. and J. Foster (2021), "A call to action: A critical review of mental health related anti-stigma campaigns", *Frontiers in Public Health*, Vol. 8, p. 990, https://doi.org/10.3389/FPUBH.2020.569539/BIBTEX. [94]

Ware, J. et al. (2002), *How to score SF-12 items: How to score version 2 of the SF-12 Health Survey*, https://www.researchgate.net/publication/291994160. [90]

Warr, D., R. Mann and T. Tacticos (2011), "Using peer-interviewing methods to explore place-based disadvantage: Dissolving the distance between suits and civilians", *International Journal of Social Research Methodology*, Vol. 14/5, pp. 337-352, https://doi.org/10.1080/13645579.2010.537527. [126]

Warwick Medical School (2021), *The Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS)*, https://warwick.ac.uk/fac/sci/med/research/platform/wemwbs/. [78]

Woodhead, C. et al. (2012), "Impact of exposure to combat during deployment to Iraq and Afghanistan on mental health by gender", *Psychological Medicine*, Vol. 42/9, pp. 1985-1996, https://doi.org/10.1017/S003329171100290X. [106]

Yamazaki, S., S. Fukuhara and J. Green (2005), "Usefulness of five-item and three-item Mental Health Inventories to screen for depressive symptoms in the general population of Japan", *Health and Quality of Life Outcomes*, Vol. 3/1, pp. 1-7, https://doi.org/10.1186/1477-7525-3-48. [39]

Notes

[1] Some screening tools contain questions relating to experienced symptoms. However, we differentiate the category of "screening tools" from that of "experienced symptoms" in that the former are set piece instruments, validated against clinical diagnoses for mental health conditions, while the former are general, non-standardised question formulations asking respondents whether, for example, they "currently experience symptoms of PTSD" or "suffer from chronic anxiety" (see Table 2.7). Refer to Chapter 2 for an extended discussion on different instrument types.

[2] This chapter covers only four composite scales of positive mental health: the SF-12, the WHO-5, MHC-SF and (S)WEMWBS. Measurement guidelines for life evaluation, affect and eudaimonic aspects of positive mental health and subjective well-being are covered in depth in (OECD, 2013[1]).

[3] By construction, screening tools with fewer items will have lower values of Cronbach's alpha. (Recall that Cronbach's alpha is a function of, among other things, the total number of items in a scale.) This again underscores the importance of weighting all facets of statistical quality together, rather than placing high importance on any single test.

[4] The two other anxiety scales against which the GAD-7 and GAD-2 were tested for convergent validity were the Beck Anxiety Inventory (BAI) and the anxiety subscale of the Symptom Checklist-90 (SCL-90).

[5] Rasch analysis uses psychometric models to analyse categorical data and identify and measure latent attitudes or characteristics.

[6] Although stigma and low levels of mental health literacy are strong drivers of non-response rates for mental health survey items, other factors – such as low levels of institutional trust, lack of motivation or sufficient time to participate, language barriers, poor health – may also contribute to low response rates (Lowthian and Lloyd, 2020[162]).

[7] Strong confidentiality assurances can reduce non-response rates for sensitive subjects (Singer, Von Thurn and Miller, 1995[58]); however, they can in fact *increase* non-response rates for non-sensitive topics, as respondents are primed to then expect threatening or sensitive questions following an in-depth data confidentiality explanation and can be put off the interview (Singer, Hippler and Schwarz, 1992[156]).

[8] The correlation between risk for depressive disorders and stigma as measured through anti-stigma indicators – the share who agree that seeking treatment for mental disorders is a sign of strength, and the share who agree that mental illness is an illness like any other – show the reverse relationship, with prevalence lower in places with less bias. However, these correlations are not significant.

[9] Cross-country and cross-group comparability are not trivial measurement issues, and some previous OECD work has dealt with the challenge of cross-country comparisons by assigning the bottom quintile of the population as at risk for mental distress, based on evidence from epidemiological studies stating that 20% of the population experiences some form of mental disorder in a given 12-month period (OECD, 2021[57]). However, this approach by definition assigns constant prevalence, which especially in the aftermath of the COVID-19 pandemic – which saw governments across the OECD struggling to deal with huge spikes in population mental distress, depression, anxiety and stress – is limiting.

[10] The geographic range where the WHO-5 has been used in surveys encompasses: Africa (Algeria, South Africa), Asia (Bangladesh, China, India, Japan, South Korea, Sri Lanka, Taiwan, Thailand), Europe (Northern, Southern, Eastern, Western and Central Europe), the Americas (Canada, the United States, Brazil, Mexico), the Middle East (Israel, Iran, Lebanon) and Oceania (Australia, New Zealand) (Topp et al., 2015[49]).

[11] These emerging methods rely heavily on the application of modern psychometric methods, such as item response theory (IRT), to improve the validity, accuracy, comparability and efficiency of mental health scales, which have in turn shown substantial promise in the advanced analysis of cross-cultural differences. Using IRT-based differential item functioning as well as the use of item anchoring or equating, new methods are able to adjust for any significant bias (Dere et al., 2015[149]; Gibbons and Skevington, 2018[141]; Vaughn-Coaxum, Mair and Weisz, 2016[155]). Similarly, new IRT models have emerged that can estimate and correct for extreme response styles more effectively than classical methods and quantify the tendency of extreme responding on a particular scale (Dowling et al., 2016[150]; Jin and Wang, 2014[152]). Some of these new methods include item banking, adapting testing and data-driven short scales and scale equating.

[12] While most international research has confirmed this rising trend of mental ill-health (OECD, 2021[98]; Santomauro et al., 2021[161]), evidence from individual countries at times show slightly different trajectories of mental health outcomes. A German study found that the prevalence of depression fell in the first year of the pandemic, but began rising by October 2020 and subsequently increased further over the course of 2021 and 2022 (Hapke et al., 2022[158]; Mauz et al., 2022[159]). An epidemiological study in Norway found that the prevalence of mental disorders decreased slightly in the early days of the pandemic (May 2020), before returning to pre-pandemic levels by September 2020 – suggesting relatively stable levels of mental disorders (Knudsen et al., 2021[160]). This mirrors findings from a meta-analysis of 65 studies from early 2020 which showed only a small average increase in mental health symptoms in March and April 2020 that had abated by July. Both studies concluded by early Q3 2020, leaving open the possibility that an extension of the research might unveil findings similar to that of Germany – little to no change in the early days of the pandemic, but rises in distress by late 2020 and 2021.

[13] These patterns also exist for physical health outcomes. A joint United States and Canada study found that self-administered respondents were more likely to report lower health-related quality-of-life (HRQoL) outcomes than did interviewer-administered telephone survey respondents (Hanmer, Hays and Fryback, 2007[142]); another study in Spain found that respondents reported better physical health outcomes, measured by the SF-36, when surveys were administered by interviewers (García et al., 2005[143]).

[14] CCHS surveys include both computer-assisted personal interviews (CAPI) and computer-assisted telephone interviews (CATI). Between 2001 and 2003, the survey changed the ratio of CATI to CAPI interviews, which allowed researchers to study how mode effects affected the comparability of CCHS data across rounds. They found differences in health indicator outcomes by mode: those interviewed in person reported higher obesity rates and were more likely to be inactive, to smoke and to report contacts with medical doctors. However, self-reported mental health showed no mode effects (St-Pierre and Béland, 2004[117]).

[15] In 2020 and 2021, many countries that still use face-to-face data collection switched to telephone surveys due to the COVID-19 pandemic. These mode shifts have not been included in the figure, as mental health outcomes in these years would be heavily influenced by the global pandemic.

[16] However, it is impossible to disentangle mode effects from the socio-political events that may have necessitated Gallup to change modes in the first place, which would be expected to exhibit an influence on underlying mental health. Many of the mode switches highlighted in this figure take place in countries that experienced significant political disruptions, or incidents of violence, that likely informed Gallup's choice to change the mode of data collection in the first place. For example, mode switches in Türkiye coincide with the 2016 attempted coup; the mode switch in 2013 in Iraq coincides with a ramping up of ISIS activity in the region; the mode switch in Libya coincides with the start of the second civil war; and so on. All of these events have a real impact on population negative affect balance and would very likely drive some of the changes shown in the figure, independently of mode effects.

[17] For example, while mental disorders are still largely perceived as shameful in Mexico, the Mexican National Comorbidity Survey interviewers experienced few refusal rates and over the course of speaking with respondents found that people were willing to open up about their mental health, often for the first time ever (Medina-Mora et al., 2008[153]).

[18] Floor effects occur when there is bunching at the lower end of the scale, whereas ceiling effects occur when there is bunching at the upper end of the scale. In Figure 3.9, the GHQ-12 shows floor effects in that most respondents fall at the lower end of the scale, which indicates they are not at any significant risk for mental distress. Because the scale focuses on those experiencing distress, it may then be less sensitive at distinguishing between individuals with higher levels of underlying positive mental health.

[19] However, some studies suggest a ceiling effect is present for (S)WEMWBS.

[20] Despite the rigor of the clinical validation process, criticisms of threshold scores remain. The cut-off scores that optimise sensitivity and specificity can differ – at times considerably – across population groups, and as a result alternatives to the use of cut-offs have been proposed (Goldberg, Oldehinkel and Ormel, 1998[151]). One such proposal is the application of stratum-specific likelihood ratios, rather than fixed thresholds, so as to allow for more detailed classification systems (Furukawa et al., 2001[144]; Furukawa and Goldberg, 1999[145]). Additionally, new findings from research into self-reported symptoms have found that the use of single sum-scores and clinical cut-offs to estimate risk for major depression may conceal important clinical insights into depression research (Fried, 2017[146]). To overcome these issues, some researchers have recommended the use of multiple depression scales to generate robust and generalisable conclusions, or the use of scales that include important non-DSM symptoms (e.g. the Symptoms of Depression Questionnaire (Pedrelli et al., 2013[147])). While it is useful to note these nuances, for a government agency measuring population mental health at a macro level – as opposed to a healthcare professional at a clinical level – there is little to suggest that the use of threshold scores is inappropriate or uninformative.

[21] Other studies have used cut-off scores ranging from 54 to 76 (Thorsen et al., 2013[130]; Hoeymans et al., 2004[154]).

[22] A respondent is deemed to be at risk for major depressive disorder if they answer "more than half the days" to either of the first two questions on the PHQ-8, and, in addition, a total of five or more of the eight items are reported as "more than half the days" (Eurostat, n.d.[74]).

[23] Lack of consistency in cut-off score usage can lead to confusion. For example, in 2019 three entities in Los Angeles provided wildly different estimates for the prevalence of mental health conditions among the homeless population. The *Los Angeles Times*, the Los Angeles Homeless Services Authority and the California Policy Lab at the University of California Los Angeles made estimates of 67%, 29% and 78%, respectively. All these came from the same dataset, with differences stemming from statistical interpretation (Smith and Oreskes, 2019[157]).

[24] Other researchers have suggested fixed cut-off points on the SWEMWBS scale: low mental well-being (having a score between 7.00 and 19.98), moderate (19.99 to 29.30) and high (29.31 to 35.00). These categories are derived from previous work on the Danish population, with low mental well-being corresponding to the bottom 15th percentile of the distribution, and high mental well-being the top 15th percentile (Santini et al., 2022[148]). Fixed cut-off scores have not been developed for the full-length WEMWBS.

[25] Care should be taken when comparing statistics on risk for major depressive disorder, or risk for depressive symptoms, coming from the PHQ-4 vs. PHQ-8 or -9. There are a number of scoring conventions for the PHQ that can lead to different prevalence estimates. Directly comparable estimates can be created by calculating risk for depression from the two individual indicators that appear in both the PHQ-4 and the PHQ-8. In this way, measures between general social and health surveys can be fully aligned, even if other (historical) health reporting has used the full set of PHQ-8 indicators to estimate depression risk prevalence.

Measuring Population Mental Health

Good mental health is a vital part of people's well-being, and the COVID-19 pandemic brought renewed attention to its importance. However, discussions so far have not focused sufficiently on how governments should best monitor it at the broader population level, and on how to consider both mental ill-health and positive mental states. This report supports national statistical offices and other data producers in collecting high-quality measures of population mental health outcomes in a more frequent, consistent and internationally harmonised manner. It documents existing measurement practice across OECD countries, discusses the advantages and limitations of available measurement tools, and recommends priority measures to adopt in household, social and health surveys. *Measuring Population Mental Health* is the first of two reports as part of an assessment of mental health and well-being in the context of the OECD's work on measuring well-being.

PRINT ISBN 978-92-64-63593-7
PDF ISBN 978-92-64-62576-1