Concurrent Validity and Reliability of Suicide Risk Assessment Instruments: A Meta-Analysis of 20 Instruments Across 27 International Cohorts

Adrian I. Campos[1, 2], Laura S. Van Velzen[3, 4], Dick J. Veltman[5, 6], Elena Pozzi[3, 4], Sonia Ambrogi[7], Elizabeth D. Ballard[8], Nerisa Banaj[7], Zeynep Başgöze[9], Sophie Bellow[10], Francesco Benedetti[11], Irene Bollettini[11], Katharina Brosch[12, 13], Erick J. Canales-Rodríguez[14, 15, 16], Emily K. Clarke-Rubright[17], Lejla Colic[18, 19, 20], Colm G. Connolly[21], Philippe Courtet[22, 23], Kathryn R. Cullen[9], Udo Dannlowski[24], Maria R. Dauvermann[10, 25], Christopher G. Davey[26], Jeremy Deverdun[27], Katharina Dohm[24], Tracy Erwin-Grabner[28], Roberto Goya-Maldonado[28], Negar Fani[29], Lydia Fortea[30, 31], Paola Fuentes-Claramonte[14, 15], Ali Saffet Gonul[32, 33, 34], Ian H. Gotlib[35], Dominik Grotegerd[24], Mathew A. Harris[36], Ben J. Harrison[37], Courtney C. Haswell[17], Emma L. Hawkins[36], Dawson Hill[9], Yoshiyuki Hirano[38, 39], Tiffany C. Ho[40, 41, 42], Fabrice Jollant[43, 44, 45, 46, 47], Tanja Jovanovic[48], Tilo Kircher[12, 13], Bonnie Klimes-Dougan[49], Emmanuelle le Bars[27, 50], Christine Lochner[51], Andrew M. McIntosh[36], Susanne Meinert[24], Yara Mekawi[29], Elisa Melloni[11], Philip Mitchell[52], Rajendra A. Morey[17, 53], Akiko Nakagawa[38, 39], Igor Nenadić[12, 13], Emilie Olié[22, 23], Fabricio Pereira[54, 55], Rachel D. Phillips[17], Fabrizio Piras[7], Sara Poletti[11], Edith Pomarol-Clotet[14, 15], Joaquim Radua[30, 56, 57], Kerry J. Ressler[29, 58], Gloria Roberts[52], Elena Rodriguez-Cano[14, 59], Matthew D. Sacchet[60], Raymond Salvador[14, 15], Anca-Larisa Sandu[61], Eiji Shimizu[38, 39], Aditya Singh[28], Gianfranco Spalletta[7, 62], J. Douglas Steele[63], Dan J. Stein[64], Frederike Stein[12, 13], Jennifer S. Stevens[29], Giana I. Teresi[35, 65], Aslihan Uyar-Demir[32], Nic J. van der Wee[66, 67], Steven J. van der Werff[66], Sanne J. H. van Rooij[29], Daniela Vecchio[7], Norma Verdolini[68], Eduard Vieta[68], Gordon D. Waiter[69], Heather Whalley[36], Sarah L. Whittle[37], Tony T. Yang[40], Carlos A. Zarate Jr.[8], Paul M. Thompson[70], Neda Jahanshad[70], Anne-Laura van Harmelen[10, 71], Hilary P. Blumberg[19, 72, 73], Lianne Schmaal[3, 4], and Miguel E. Rentería[1, 2]

[1] Department of Genetics & Computational Biology, QIMR Berghofer Medical Research Institute
[2] Faculty of Medicine, School of Biomedical Sciences, The University of Queensland
[3] Orygen, Parkville, Victoria, Australia
[4] Centre for Youth Mental Health, University of Melbourne
[5] Department of Psychiatry, Amsterdam University Medical Center, Vrije Universiteit Medical Center
[6] Amsterdam Neuroscience, Amsterdam, the Netherlands
[7] Department of Clinical and Behavioral Neurology, Laboratory of Neuropsychiatry, IRCCS Santa Lucia Foundation, Rome, Italy
[8] Experimental Therapeutics and Pathophysiology Branch, National Institute of Mental Health, Bethesda, Maryland, United States
[9] Department of Psychiatry and Behavioral Sciences, University of Minnesota
[10] Department of Psychiatry, University of Cambridge
[11] IRCCS San Raffaele Scientific Institute, Vita-Salute San Raffaele University
[12] Department of Psychiatry and Psychotherapy, University of Marburg
[13] Center for Mind Brain and Behavior, University of Marburg
[14] FIDMAG Germanes Hospitalaries Research Foundation, Barcelona, Spain
[15] Centro de Investigación Biomédica En Red de Salud Mental, Barcelona, Spain
[16] Signal Processing Laboratory (LTS5), École polytechnique fédérale de Lausanne, Lausanne, Switzerland
[17] Brain Imaging and Analysis Center, Duke University School of Medicine
[18] Department of Psychiatry and Psychotherapy, University Hospital Jena
[19] Department of Psychiatry, Yale School of Medicine, Yale University
[20] German Center for Mental Health, Halle-Jena-Magdeburg, Germany
[21] Magnetic Resonance Imaging Facility and Department of Biomedical Sciences, Florida State University
[22] Department of Emergency Psychiatry and Acute Care, Lapeyronie Hospital, Centre Hospitalier Universitaire Montpellier, Montpellier, France
[23] Institut de Génomique Fonctionnelle, University of Montpellier, Centre national de la recherche scientifique, Institut national de la santé et de la recherche médicale
[24] Institute for Translational Psychiatry, University of Münster
[25] Department of Forensic and Neurodevelopmental Sciences, King's College London
[26] Department of Psychiatry, The University of Melbourne & Melbourne Health

[27] Institut d'Imagerie Fonctionnelle Humaine, I2FH, Montpellier University Hospital Center, Gui de Chauliac Hospital, University of Montpellier
[28] Department of Psychiatry and Psychotherapy, Laboratory of Systems Neuroscience and Imaging in Psychiatry (SNIP-Lab), University Medical Center Goettingen, Georg-August University
[29] Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine
[30] Institut d'Investigacions Biomèdiques August Pi i Sunyer, Centro de Investigación Biomédica en Red en Salud Mental, Barcelona, Spain
[31] Institute of Neurosciences, University of Barcelona
[32] Department of Psychiatry, Laboratory for Standardization of Computational Anatomy Techniques for Cognitive and Behavioral Sciences, School of Medicine, Ege University
[33] Department of Neuroscience, Institute of Health Sciences, Ege University
[34] Department of Psychiatry and Behavioral Sciences, School of Medicine, Mercer University
[35] Department of Psychology, Stanford University
[36] Division of Psychiatry, Centre for Clinical Brain Sciences, University of Edinburgh
[37] Department of Psychiatry, Melbourne Neuropsychiatry Centre, The University of Melbourne
[38] Research Center for Child Mental Development, Chiba University
[39] United Graduate School of Child Development, Osaka University, Kanazawa University, Hamamatsu University School of Medicine, Chiba University and University of Fukui
[40] Department of Psychiatry and Behavioral Sciences, Division of Child and Adolescent Psychiatry, Weill Institute for Neurosciences, University of California, San Francisco (UCSF) School of Medicine
[41] Department of Psychology, Stanford University
[42] Department of Psychiatry and Behavioral Sciences and Weill Institute for Neurosciences, University of California, San Francisco
[43] Groupe Hospitalier Universitaire Paris Psychiatrie et Neurosciences, Paris, France
[44] McGill Group for Suicide Studies, McGill University
[45] Department of Psychiatrie, CHU Nîmes
[46] Universitätsklinikum Jena
[47] Department of Psychiatry, Université de Paris
[48] Department of Psychiatry and Behavioral Neurosciences, Wayne State University School of Medicine
[49] Department of Psychology, University of Minnesota
[50] Department of Neuroradiology, Montpellier University Hospital, Gui de Chauliac Hospital
[51] Department of Psychiatry, South African Medical Research Council Unit on Risk & Resilience in Mental Disorders, Stellenbosch University
[52] School of Psychiatry, University of New South Wales
[53] Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine
[54] Department of Radiology and Psychiatry, University Hospital Center of Nîmes
[55] Mathématiques, Informatique, Physique et Applications, University of Nîmes
[56] Department of Psychosis Studies, Institute of Psychiatry, Psychology and Neuroscience, King's College London
[57] Department of Clinical Neuroscience, Centre for Psychiatric Research and Education, Karolinska Institutet, Stockholm, Sweden
[58] McLean Hospital, Harvard Medical School, Boston, Massachusetts, United States
[59] Benito Menni CASM, Sant Boi de Llobregat, Barcelona, Spain
[60] Center for Depression, Anxiety, and Stress Research, McLean Hospital, Harvard Medical School, Belmont, Massachusetts, United States
[61] Aberdeen Biomedical Imaging Centre, Institute of Medical Sciences, University of Aberdeen
[62] Division of Neuropsychiatry, Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine
[63] Division of Imaging Science and Technology, Medical School, University of Dundee
[64] Department of Psychiatry & Neuroscience Institute, South African Medical Research Council Unit on Risk & Resilience in Mental Disorders, University of Cape Town
[65] Department of Psychology, University of Pittsburgh
[66] Department of Psychiatry, Leiden University Medical Center
[67] Leiden Institute for Brain and Cognition, Leiden University Medical Center
[68] Hospital Clinic, Institute of Neuroscience, IDIBAPS, Centro de Investigación Biomédica En Red of Mental Health, University of Barcelona
[69] Aberdeen Biomedical Imaging Centre, Institute of Medical Sciences, University of Aberdeen
[70] Mark and Mary Stevens Neuroimaging and Informatics Institute, Imaging Genetics Center, Keck School of Medicine, University of Southern California, Los Angeles
[71] Social Security and Resilience Programme, Education and Child Studies, Leiden University
[72] Department of Radiology and Biomedical Imaging, Yale School of Medicine, Yale University
[73] Child Study Center, Yale School of Medicine, Yale University

Objective: A major limitation of current suicide research is the lack of power to identify robust correlates of suicidal thoughts or behavior. Variation in suicide risk assessment instruments used across cohorts may

represent a limitation to pooling data in international consortia. *Method:* Here, we examine this issue through two approaches: (a) an extensive literature search on the reliability and concurrent validity of the most commonly used instruments and (b) by pooling data ($N \sim 6,000$ participants) from cohorts from the Enhancing NeuroImaging Genetics Through Meta-Analysis (ENIGMA) Major Depressive Disorder and ENIGMA–Suicidal Thoughts and Behaviour working groups, to assess the concurrent validity of instruments currently used for assessing suicidal thoughts or behavior. *Results:* We observed moderate-to-high correlations between measures, consistent with the wide range (κ range: 0.15–0.97; r range: 0.21–0.94) reported in the literature. Two common multi-item instruments, the Columbia Suicide Severity Rating Scale and the Beck Scale for Suicidal Ideation were highly correlated with each other ($r = 0.83$). Sensitivity analyses identified sources of heterogeneity such as the time frame of the instrument and whether it relies on self-report or a clinical interview. Finally, construct-specific analyses suggest that suicide ideation items from common psychiatric questionnaires are most concordant with the suicide ideation construct of multi-item instruments. *Conclusions:* Our findings suggest that multi-item instruments provide valuable information on different aspects of suicidal thoughts or behavior but share a modest core factor with single suicidal ideation items. Retrospective, multisite collaborations including distinct instruments should be feasible provided they harmonize across instruments or focus on specific constructs of suicidality.

Key Points
Question: To inform future suicide research in multisite international consortia, it is important to examine how different suicide measures relate to each other and whether they can be used interchangeably. *Findings:* Findings suggest that detailed instruments (such as the Columbia Suicide Severity Rating Scale and Beck Scale for Suicidal Ideation) provide valuable information on suicidal thoughts and behavior and share a core factor with items on suicidal ideation from depression severity rating scale (such as the Hamilton Depression Rating Scale or the Beck Depression Inventory). *Importance:* Results from international collaborations can mitigate biases by harmonizing distinct suicide risk assessment instruments and items related to suicide. *Next Steps:* Pooling data within international suicide research consortia may reveal novel clinical, biological, and cognitive correlates of suicidal thoughts and/or behavior.

Adrian I. Campos https://orcid.org/0000-0003-3468-8619
Laura S. Van Velzen https://orcid.org/0000-0002-6003-5227
Elena Pozzi https://orcid.org/0000-0001-8360-5571
Sonia Ambrogi https://orcid.org/0000-0002-2897-358X
Nerisa Banaj https://orcid.org/0000-0002-6317-8549
Zeynep Başgöze https://orcid.org/0000-0002-0306-4849
Irene Bollettini https://orcid.org/0000-0003-3666-3538
Katharina Brosch https://orcid.org/0000-0002-0526-8095
Erick J. Canales-Rodríguez https://orcid.org/0000-0001-6421-2633
Emily K. Clarke-Rubright https://orcid.org/0000-0003-3553-2990
Lejla Colic https://orcid.org/0000-0002-8809-4034
Colm G. Connolly https://orcid.org/0000-0002-4051-3682
Philippe Courtet https://orcid.org/0000-0002-6519-8586
Kathryn R. Cullen https://orcid.org/0000-0001-9631-3770
Maria R. Dauvermann https://orcid.org/0000-0002-2873-8512
Christopher G. Davey https://orcid.org/0000-0003-1431-3852
Jeremy Deverdun https://orcid.org/0000-0002-7083-2940
Negar Fani https://orcid.org/0000-0002-7720-252X
Lydia Fortea https://orcid.org/0000-0001-6778-2316
Paola Fuentes-Claramonte https://orcid.org/0000-0002-1428-7976
Ian H. Gotlib https://orcid.org/0000-0002-3622-3199
Dominik Grotegerd https://orcid.org/0000-0001-8718-0128
Courtney C. Haswell https://orcid.org/0000-0003-0562-9855
Emma L. Hawkins https://orcid.org/0000-0002-0241-7251
Yoshiyuki Hirano https://orcid.org/0000-0003-3844-3061
Tiffany C. Ho https://orcid.org/0000-0002-4500-6364
Fabrice Jollant https://orcid.org/0000-0001-5809-4503
Tanja Jovanovic https://orcid.org/0000-0002-6456-0053
Emmanuelle le Bars https://orcid.org/0000-0002-3927-9868

Yara Mekawi https://orcid.org/0000-0002-9181-2324
Elisa Melloni https://orcid.org/0000-0002-6742-8373
Fabricio Pereira https://orcid.org/0000-0003-3317-8756
Rachel D. Phillips https://orcid.org/0000-0001-9958-1703
Fabrizio Piras https://orcid.org/0000-0003-3566-5494
Sara Poletti https://orcid.org/0000-0001-9594-0246
Joaquim Radua https://orcid.org/0000-0003-1240-5438
Kerry J. Ressler https://orcid.org/0000-0002-5158-1103
Raymond Salvador https://orcid.org/0000-0001-5557-1562
Anca-Larisa Sandu https://orcid.org/0000-0002-5555-4129
Aditya Singh https://orcid.org/0000-0002-0539-7129
Gianfranco Spalletta https://orcid.org/0000-0002-7432-4249
J. Douglas Steele https://orcid.org/0000-0002-9822-8753
Dan J. Stein https://orcid.org/0000-0001-7218-7810
Frederike Stein https://orcid.org/0000-0002-5052-6022
Jennifer S. Stevens https://orcid.org/0000-0003-4674-0314
Giana I. Teresi https://orcid.org/0000-0002-9674-3313
Norma Verdolini https://orcid.org/0000-0001-9488-2881
Heather Whalley https://orcid.org/0000-0002-4505-8869
Neda Jahanshad https://orcid.org/0000-0003-4401-8950
Hilary P. Blumberg https://orcid.org/0000-0002-6363-4698
Lianne Schmaal https://orcid.org/0000-0001-9822-048X
Miguel E. Rentería https://orcid.org/0000-0003-4626-7248
Adrian I. Campos and Laura S. Van Velzen contributed equally to this work and share first authorship.

curation, writing of original draft, and writing of review and editing. Sonia Ambrogi played supporting role in data curation, writing of original draft, and writing of review and editing. Elizabeth D. Ballard played supporting role in data curation, writing of original draft, and writing of review and editing. Nerisa Banaj played supporting role in data curation, writing of original draft, and writing of review and editing. Zeynep Başgöze played supporting role in data curation, writing of original draft, and writing of review and editing. Sophie Bellow played supporting role in data curation, writing of original draft, and writing of review and editing. Francesco Benedetti played supporting role in data curation, writing of original draft, and writing of review and editing. Irene Bollettini played supporting role in data curation, writing of original draft, and writing of review and editing. Katharina Brosch played supporting role in data curation, writing of original draft, and writing of review and editing. Erick J. Canales-Rodríguez played supporting role in data curation, writing of original draft, and writing of review and editing. Emily K. Clarke-Rubright played supporting role in data curation, writing of original draft, and writing of review and editing. Lejla Colic played supporting role in data curation, writing of original draft, and writing of review and editing. Colm G. Connolly played supporting role in data curation, writing of original draft, and writing of review and editing. Philippe Courtet played supporting role in data curation, writing of original draft, and writing of review and editing. Kathryn R. Cullen played supporting role in data curation, writing of original draft, and writing of review and editing. Udo Dannlowski played supporting role in data curation, writing of original draft, and writing of review and editing. Maria R. Dauvermann played supporting role in data curation, writing of original draft, and writing of review and editing. Christopher G. Davey played supporting role in data curation, writing of original draft, and writing of review and editing. Jeremy Deverdun played supporting role in data curation, writing of original draft, and writing of review and editing. Katharina Dohm played supporting role in data curation, writing of original draft, and writing of review and editing. Tracy Erwin-Grabner played supporting role in data curation, writing of original draft, and writing of review and editing. Roberto Goya-Maldonado played supporting role in data curation, writing of original draft, and writing of review and editing. Negar Fani played supporting role in data curation, writing of original draft, and writing of review and editing. Lydia Fortea played supporting role in data curation, writing of original draft, and writing of review and editing. Paola Fuentes-Claramonte played supporting role in data curation, writing of original draft, and writing of review and editing. Ali Saffet Gonul played supporting role in data curation, writing of original draft, and writing of review and editing. Ian H. Gotlib played supporting role in data curation, writing of original draft, and writing of review and editing. Dominik Grotegerd played supporting role in data curation, writing of original draft, and writing of review and editing. Mathew A. Harris played supporting role in data curation, writing of original draft, and writing of review and editing. Ben J. Harrison played supporting role in data curation, writing of original draft, and writing of review and editing. Courtney C. Haswell played supporting role in data curation, writing of original draft, and writing of review and editing. Emma L. Hawkins played supporting role in data curation, writing of original draft, and writing of review and editing. Dawson Hill played supporting role in data curation, writing of original draft, and writing of review and editing. Yoshiyuki Hirano played supporting role in data curation, writing of original draft, and writing of review and editing. Tiffany C. Ho played supporting role in data curation, writing of original draft, and writing of review and editing. Fabrice Jollant played supporting role in data curation, writing of original draft, and writing of review and editing. Tanja Jovanovic played supporting role in data curation, writing of original draft, and writing of review and editing. Tilo Kircher played supporting role in data curation, writing of original draft, and writing of review and editing. Bonnie Klimes-Dougan played supporting role in data curation, writing of original draft, and writing of review and editing. Emmanuelle le Bars played supporting role in data curation, writing of original draft, and writing of review and editing. Christine Lochner played supporting role in data curation,

writing of original draft, and writing of review and editing. Andrew M. McIntosh played supporting role in data curation, writing of original draft, and writing of review and editing. Susanne Meinert played supporting role in data curation, writing of original draft, and writing of review and editing. Yara Mekawi played supporting role in data curation, writing of original draft, and writing of review and editing. Elisa Melloni played supporting role in data curation, writing of original draft, and writing of review and editing. Philip Mitchell played supporting role in data curation, writing of original draft, and writing of review and editing. Rajendra A. Morey played supporting role in data curation, writing of original draft, and writing of review and editing. Akiko Nakagawa played supporting role in data curation, writing of original draft, and writing of review and editing. Igor Nenadić played supporting role in data curation, writing of original draft, and writing of review and editing. Emilie Olié played supporting role in data curation, writing of original draft, and writing of review and editing. Fabricio Pereira played supporting role in data curation, writing of original draft, and writing of review and editing. Rachel D. Phillips played supporting role in data curation, writing of original draft, and writing of review and editing. Fabrizio Piras played supporting role in data curation, writing of original draft, and writing of review and editing. Sara Poletti played supporting role in data curation, writing of original draft, and writing of review and editing. Edith Pomarol-Clotet played supporting role in data curation, writing of original draft, and writing of review and editing. Joaquim Radua played supporting role in data curation, writing of original draft, and writing of review and editing. Kerry J. Ressler played supporting role in data curation, writing of original draft, and writing of review and editing. Gloria Roberts played supporting role in data curation, writing of original draft, and writing of review and editing. Elena Rodriguez-Cano played supporting role in data curation, writing of original draft, and writing of review and editing. Matthew D. Sacchet played supporting role in data curation, writing of original draft, and writing of review and editing. Raymond Salvador played supporting role in data curation, writing of original draft, and writing of review and editing. Anca-Larisa Sandu played supporting role in data curation, writing of original draft, and writing of review and editing. Eiji Shimizu played supporting role in data curation, writing of original draft, and writing of review and editing. Aditya Singh played supporting role in data curation, writing of original draft, and writing of review and editing. Gianfranco Spalletta played supporting role in data curation, writing of original draft, and writing of review and editing. J. Douglas Steele played supporting role in data curation, writing of original draft, and writing of review and editing. Dan J. Stein played supporting role in data curation, writing of original draft, and writing of review and editing. Frederike Stein played supporting role in data curation, writing of original draft, and writing of review and editing. Jennifer S. Stevens played supporting role in data curation, writing of original draft, and writing of review and editing. Giana I. Teresi played supporting role in data curation, writing of original draft, and writing of review and editing. Aslihan Uyar-Demir played supporting role in data curation, writing of original draft, and writing of review and editing. Nic J. van der Wee played supporting role in data curation, writing of original draft, and writing of review and editing. Steven J. van der Werff played supporting role in data curation, writing of original draft, and writing of review and editing. Sanne J. H. van Rooij played supporting role in data curation, writing of original draft, and writing of review and editing. Daniela Vecchio played supporting role in data curation, writing of original draft, and writing of review and editing. Norma Verdolini played supporting role in data curation, writing of original draft, and writing of review and editing. Eduard Vieta played supporting role in data curation, writing of original draft, and writing of review and editing. Gordon D. Waiter played supporting role in data curation, writing of original draft, and writing of review and editing. Heather Whalley played supporting role in data curation, writing of original draft, and writing of review and editing. Sarah L. Whittle played supporting role in data curation, writing of original draft, and writing of review and editing. Tony T. Yang played supporting role in data

Suicide is a leading cause of death worldwide, with an estimated 800,000 deaths by suicide occurring annually, or one person dying by suicide every 40 s (World Health Organization [WHO], 2014). Despite national and international efforts to prevent suicide, suicide rates continue to rise in some countries around the world (Alicandro et al., 2019).

To better understand and identify demographic, environmental, psychological, cognitive, and neurobiological factors associated with suicidal thoughts and behavior (STB), we need large samples, as individual factors most likely explain a small proportion of complex phenotypes as suicidal thoughts or behaviors. Large and diverse samples additionally provide the opportunity to study the heterogeneity in associated factors by identifying subgroups and moderation effects. Large-scale international collaborations in consortia for suicide research may provide an important step forward to increase our understanding of STB.

Examples of these consortia include the International Suicide Genetics Consortium (Mullins et al., 2022) and the Enhancing NeuroImaging Genetics through Meta-Analysis Suicidal Thoughts and Behaviour (ENIGMA-STB) consortium (van Velzen et al., 2022). The aim of ENIGMA-STB is to study the neural correlates of STB, by bringing together research groups around the world that have collected both neuroimaging data and assessed STB in individuals with and without mental disorders. These groups use standardized protocols to process their neuroimaging data and then pool data in analyses that have increased statistical power to detect relevant associations.

While these large-scale collaborations have many strengths, it has been challenging to harmonize the different instruments employed to assess STBs across cohorts, and the validity of the findings will depend on the quality of STB measure harmonization. For instance, in our recent large-scale analysis of cortical morphology across 18 research groups within the ENIGMA Major Depressive Disorder (ENIGMA-MDD) consortium (Campos et al., 2020), suicide attempt was assessed using 19 different measures, including single items on STBs from depression severity questionnaires, items from clinical interviews, in addition to items from comprehensive instruments specifically focused on STBs such as the Columbia Suicide Severity Rating Scale (C-SSRS; Posner et al., 2011).

To inform future suicide research in international consortia, it is important to examine how these different suicide measures relate to each other and whether they can be used interchangeably. Therefore, the aim of this study was to examine the correlations between the 20 different assessment instruments for STBs used across 27 ENIGMA cohorts. In the first part of this report, we provide an overview of the literature on reliability and validity of these commonly used measures to assess STBs and the associations between these measures (concurrent validity). In the second part, we present findings from a meta-analysis performed using the cross-sectional measures, including multi-item and single-item measures, collected across 27 cohorts within the ENIGMA-MDD and ENIGMA-STB working groups.

Method

Literature Search

A literature search was conducted in PubMed (https://pubmed.ncbi.nlm.nih.gov) for articles published before December 2021, using the following search terms: suicid* AND (questionnaire* OR interview OR measures) AND (validity OR convergent validity OR discriminant validity OR reliability OR psychometric*), using "English" and "Human" as additional filters.

Two thousand five hundred forty-nine abstracts were screened by investigator Laura S. Van Velzen to identify studies which used psychometric measures to assess suicidal ideation and/or suicidal behavior that were also collected by the ENIGMA research groups. These measures included the Beck Depression Inventory (BDI) suicidal ideation item (Beck et al., 1961), Scale for Suicidal Ideation (SSI; Beck et al., 1979, 1988), Children's Depression Rating Scale (CDRS) suicidal ideation item (Poznanski & Mokros, 1996), Composite International Diagnostic Interview (CIDI) items on suicidal ideation and behavior (WHO, 1997), C-SSRS (Posner et al., 2011), Diagnostic Interview for Genetics Studies (DIGS) items on suicidal ideation and behavior (Nurnberger et al., 1994), Hamilton Depression Rating Scale (HAM-D) item on suicidal ideation (Hamilton, 1960), Inventory of Depression and Anxiety Symptoms (IDAS-II) suicide subscale (Watson et al., 2012), (Quick) Inventory of Depressive Symptomatology (IDS/QIDS) suicidal ideation item (Rush et al., 1986), Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS) suicide items (Kaufman et al., 1997), Montgomery–Asberg Depression Rating Scale (MADRS) suicidal ideation item (Montgomery & Asberg, 1979), Mini International Neuropsychiatric Interview (MINI) suicidality module (Sheehan et al., 1998), Revised Children's Anxiety and Depression Scale (RCADS) suicidal ideation item (Chorpita et al., 2000), Structured Clinical Interview for *Diagnostic and Statistical Manual of Mental* Disorders (SCID) suicide questions (First, 1997), Suicidal Ideation Questionnaire (SIQ; Reynolds, 1987), Beck's Suicide Intent Scale (SIS; Beck et al., 1974), Self-Injurious Thoughts and Behaviors Interview (SITBI; Nock et al., 2007), Suicide Score Scale (SSS; Innamorati et al., 2008), Youth Self-Report suicide item (YSR; Achenbach et al., 1991), SIQ–Junior (SIQ-JR; Reynolds, 1987).

A total of 225 studies were identified and screened for information on the reliability (interrater reliability, internal consistency, or test–retest reliability) or validity (correlation with an established instrument, e.g., concurrent validity or predictive validity) of those

measures. For concurrent validity, we included only associations between measures that were collected by the ENIGMA working groups and mentioned above. Additional studies were identified by cross-referencing.

Data Dimensionality Reduction Strategy

Our study comprised both multiple-item and single-item suicide risk assessment instruments. Importantly, only cross-sectional data were included. Single-item instruments such as questions from depression severity rating scales or psychiatric interviews normally assess recent suicidal ideation. Multi-item instruments typically capture other dimensions such as control over suicidal thoughts, protective factors, and, in the case of suicide attempt, degree of intent to die. By extracting common factor scores for the multi-item instruments, we are able to obtain a score of the underlying suicidal risk being measured by the instruments while reducing the need to adjust for slight differences between versions. The choice of dimensionality reduction approach, common factor scores using full-information maximum likelihood (FIML), was motivated by two reasons: (a) this approach deals with missing data, which is common in these questionnaires, using FIML and (b) we obtain a single factor score capturing the main liability measured by the instrument, as opposed to other approaches (e.g., principal component analysis) that require nonmissing data and output several new variables. Data dimensionality reduction for the SSI, KSADS, IDAS, SIQ, DIGS, C-SSRS, MINI, and SSS was performed by extracting common factor scores using FIML factor analysis as implemented in the structural equation modeling package *umx* in R (Bates, 2018; Bates et al., 2016). Notably, dimensionality reduction included all items available for each cohort (with any variance) of the multi-item instruments. For instance, the C-SSRS instrument includes items on severity of suicidal ideation, intensity of suicidal ideation, and suicidal behaviors. We did not focus on a specific section of these items in our main analyses.

Main Analysis on ENIGMA Data

We implemented a sample size–weighted meta-analysis of correlations between suicide risk assessment instruments across 27 international cohorts from the ENIGMA-MDD and ENIGMA-STB working groups. Cohorts shared de-identified individual-level response data to STBs assessment instruments or items on STBs from depression symptom severity questionnaires or clinical interviews. All participants provided informed consent, and all projects were approved by their respective relevant ethics committees. Our initial analysis consisted of three steps: (a) data dimensionality reduction for multi-item suicide risk assessment instruments (see above); (b) within-cohort unadjusted correlations for all possible pairs of instruments; and (c) a sample size–weighted meta-analysis, combining the correlation coefficients for pairs of instruments for which data were available across multiple cohorts while accounting for cohort sample size differences. Within-cohort correlations and the sample size–weighted meta-analysis were calculated in python using the *scipy* (Virtanen et al., 2020), *numpy* (Harris et al., 2020), and *pandas* libraries. Data were visualized using undirected graphs with varying node and edge sizes according to the number of cohorts and cohort pairs supporting each correlation. These were generated from the data using python and the *networkx* library (Hagberg et al., 2008). We analyzed data from two working groups of the ENIGMA

consortium, including 22 instruments across 27 cohorts worldwide. Individual-level responses for 6,716 participants were included in our study (Table 1). Participants were included across a range of diagnoses including MDD, psychotic disorders, anxiety disorders, obsessive–compulsive disorders, posttraumatic stress disorder, and bipolar disorder, along with data from healthy controls.

Sensitivity Analyses

We performed a series of sensitivity analyses to identify sources of heterogeneity that could explain our results. These analyses consisted of repeating our framework but on subsets of the data. We mainly assessed (i) whether removing items measuring protective factors affected our results; (ii) compared lifetime and recent suicide risk assessment as well as (iii) interview-based and self-report suicide risk assessment; and (iv) perform construct-specific analyses to assess the degree of concordance of construct factors (i.e., suicide attempt, suicide ideation, and suicide intent) derived from items in multi-item instruments and whether single-item instruments show a higher concordance with a specific construct. For Aim iv, items of multi-item instruments were mapped to one of three possible constructs, suicide ideation, suicide attempt, and suicide intent (Supplemental Table S1). Dimensionality reduction was then carried out for all the items mapped to each of these constructs, and the analyses were repeated using these *construct factors* instead of the previous common factors. The following constructs were assessed: suicide intent—defined as the seriousness or intensity of the wish of a person to end their life; suicidal ideation—defined as thoughts, ideas, and the desire to die by suicide, which can vary in severity and intensity; and suicide attempt—defined as a behavior or act of self-harm related to an individual intentionally seeking to end their life.

Results

Literature Review

An overview of the reliability and validity measures for the different suicide scales and items derived from our literature review is presented in Supplemental Tables S2 and S3, respectively. Information on the reliability (interrater reliability, internal consistency, and test–retest reliability) and validity (concurrent and predictive validity) of these measures was most often available for instruments specifically focused on STBs (e.g., SIS, SSI, C-SSRS), followed by items on suicidal ideation from questionnaires assessing the severity of depressive symptoms (e.g., BDI, MADRS, HAM-D). No reliability or validity measures were identified for suicide questions from diagnostic interviews (e.g., CIDI, SCID). The lowest concurrent validity measure identified was between the SIS and C-SSRS scales. Overall, mostly moderate-to-high correlation or concurrent validity scores (κ range: 0.15–0.97; r range: 0.21–0.94) between instruments were identified (Figure 1a). When examining the concurrent validity between self-report measures, the correlation ranged between 0.41 and 0.69, and κ varied between 0.15 and 0.45, while for clinician-administered measures, the correlation varied between 0.21 and 0.94 and the κ varied between 0.42 and 0.97 (Figure 1a).

With regard to reliability, as expected, the strongest test–retest reliability was seen for instruments that assess lifetime or longer term suicidal ideation or behavior (SIQ, SIQ-JR; $r = 0.72$–0.91 and 0.89, respectively), compared to varying test–retest reliability in measures that assess short-term suicidal ideation (such as the BDI or

Table 1

Cohorts Included, Sample Size, and Instruments Used to Assess Suicidal Thoughts and/or Behaviors

Cohort name	Sample size	Instruments
AFFDIS	29	Hamilton Depression Rating Scale (past week), Beck Depression Inventory (past 2 weeks), Montgomery–Asberg Depression Rating Scale (past week)
Chiba University	117	Beck Depression Inventory-I (past week), Beck Depression Inventory-II (past 2 weeks), Mini International Neuropsychiatric Interview suicidality module (past month, Question 6 refers to lifetime attempt)
Duke/Durham VA	190	Beck Depression Inventory (past 2 weeks), Beck Scale for Suicidal Ideation (past week)
EPISCA (Leiden adolescents)	71	Children's Depression Inventory (past 2 weeks), Youth Self-Report (lifetime), Revised Children's Anxiety and Depression Scale (lifetime)
ETPB-STB	60	Hamilton Depression Rating Scale (past week), Scale for Suicidal Ideation (past week), Beck Depression Inventory (past week), Montgomery–Asberg Depression Rating Scale (past week)
FIDMAG	284	Hamilton Depression Rating Scale (past week), Montgomery–Asberg Depression Rating Scale (past week)
FOR 2107 Muenster	424	Hamilton Depression Rating Scale (past week), Beck Depression Inventory (past week)
FOR 2107 Marburg	792	Hamilton Depression Rating Scale (past week), Beck Depression Inventory (past week)
Grady Trauma Project Emory University	123	Beck Depression Inventory (past 2 weeks), Columbia Suicide Severity Rating Scale (past month)
The University of Melbourne	283	Quick Inventory of Depressive Symptoms (past week), Suicidal Ideation Questionnaire (lifetime), Columbia Suicide Severity Rating Scale (past month and lifetime), Montgomery–Asberg Depression Rating Scale (past week)
University of Minnesota Adolescent MDD	110	Children's Depression Rating Scale (past week), Beck Depression Inventory (past 2 weeks), Kiddie Schedule for Affective Disorders and Schizophrenia (lifetime), Inventory of Depression and Anxiety Symptoms (past 2 weeks)
CHU Montpellier BICS study	66	Inventory of Depressive Symptoms—Clinician rated (past week), Beck Depression Inventory (current), Columbia Suicide Severity Rating Scale (lifetime and last month)
CHU Montpellier IMPACT study	40	Quick Inventory of Depressive Symptoms (past week), Inventory of Depressive Symptoms—Clinician rated (past week), Columbia Suicide Severity Rating Scale (last month)
CHU Montpellier Servier Study	120	Hamilton Depression Rating Scale (past week), Beck Depression Inventory (past week), Scale for Suicidal Ideation (past week)
McGill University	103	Beck Depression Inventory (past 2 weeks), Suicide Intent Scale (most recent attempt and most severe attempt), Scale for Suicidal Ideation (today and in the past 2 weeks), Quick Inventory of Depressive Symptoms (past week), Montgomery–Asberg Depression Rating Scale (past week), Hamilton Depression Rating Scale (past week), Columbia Suicide Severity Rating Scale (past month and lifetime)
Moral Dilemma	62	Quick Inventory of Depressive Symptoms (past week), Montgomery–Asberg Depression Rating Scale (past week)
Muenster Neuroimaging Cohort	1,064	Hamilton Depression Rating Scale (past week), Beck Depression Inventory (past week), Montgomery–Asberg Depression Rating Scale (past week)
Fondazione Santa Lucia	288	Suicide Score Scale (past year), Mini International Neuropsychiatric Interview suicidality module (past month)
San Raffaele Hospital	447	Hamilton Depression Rating Scale (past week), Beck Depression Inventory (past 2 weeks), Beck Scale for Suicidal Ideation (day of assessment)
SoCAT	179	Hamilton Depression Inventory (past week), Beck Depression Inventory (past 2 weeks)
South Africa	117	Diagnostic Interview for Genetic Studies (lifetime), Montgomery–Asberg Depression Rating Scale (past week)
Stanford University adolescent MDD TIGER	49	Columbia Suicide Severity Rating Scale (lifetime and current [ideation in the past week, attempt in the past month]), Self-Injurious Thoughts and Behaviors Interview (lifetime)
Stanford University AGG/FAA	56	Structured Clinical Interview for *DSM* Disorders (lifetime), Hamilton Depression Rating Scale (past week)
STRADL	1,188	Structured Clinical Interview for *DSM* Disorders (past month), Quick Inventory of Depressive Symptoms (past week)
Sydney Bipolar Risk Study	225	Kiddie Schedule for Affective Disorders and Schizophrenia (lifetime), Diagnostic Interview for Genetic Studies (lifetime), Montgomery–Asberg Depression Rating Scale (lifetime)
UCSF Adolescent MDD	161	Beck Depression Inventory (2 weeks), Columbia Suicide Severity Scale (lifetime and current [ideation in the past week, attempt in the past 2 weeks])
Yale School of Medicine	178	Beck Scale for Suicidal Ideation (past month and lifetime), Suicide Intent Scale (most recent attempt, most lethal attempt), Columbia Suicide Severity Rating Scale (lifetime and past month), Hamilton Depression Rating Scale (past week), Children's Depression Rating Scale (past week)

Note. VA = Veteran Affairs; STB = Suicidal Thoughts and Behaviour; CHU = Centre Hospitalier Universitaire; *DSM* = *Diagnostic and Statistical Manual of Mental Disorders*; UCSF = University of California, San Francisco; MDD = Major Depressive Disorder.

Figure 1

Overview of Results
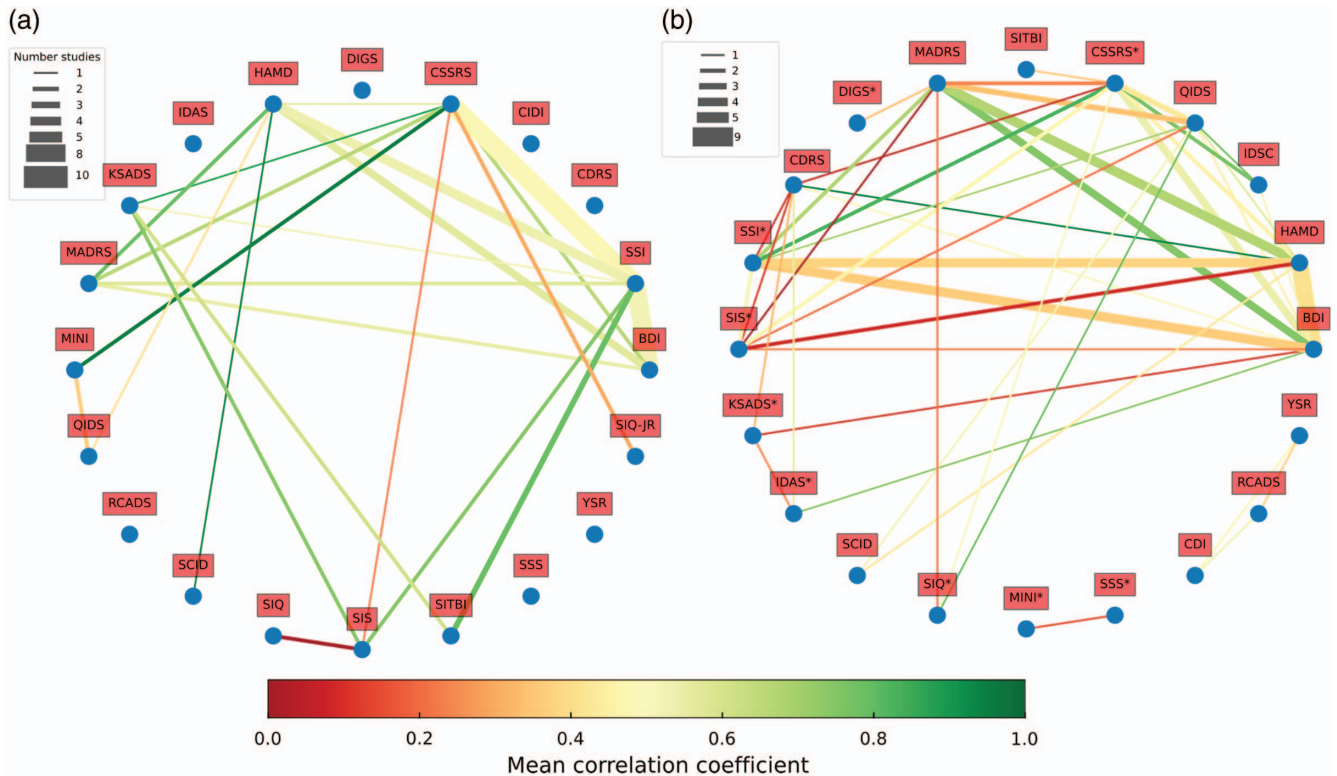
Note. (a) Literature review results. Reported instrument correlations are shown using an undirected graph. Nodes represent the instruments studied. Edges are colored based on the average correlation across cohorts, edge width increases with the number of studies from which the correlations were extracted. (b) ENIGMA correlation results. Each node represents one of the instruments included in the study. Edge color represents the sample size–weighted average correlation coefficient between two instruments. The thickness of the edge increases with the number of cohorts contributing to estimate the correlation. Generally speaking, the thicker the edge the more confidence in the correlation estimate. (c) Overview of the instruments used on the enigma analyses and the amount of overlap between them. BDI = Beck Depression Inventory suicidal ideation item; SSI = Scale for Suicidal Ideation; CDRS = Children's Depression Rating Scale suicidal ideation item; CIDI = Composite International Diagnostic Interview items on suicidal ideation and behavior; C-SSRS = Columbia Suicide Severity Rating Scale; DIGS = Diagnostic Interview for Genetics Studies items on suicidal ideation and behaviour; HAM-D = Hamilton Depression Rating Scale item on suicidal ideation; IDAS-II = Inventory of Depression and Anxiety Symptoms suicide subscale; CSSRS = Columbia Suicide Severity Rating Scale; QIDS = Quick Inventory of Depressive Symptomatology suicidal ideation item; KSADS = Kiddie Schedule for Affective Disorders and Schizophrenia suicide items; MADRS = Montgomery–Asberg Depression Rating Scale suicidal ideation item; MINI = Mini International Neuropsychiatric Interview suicidality module; RCADS = Revised Children's Anxiety and Depression Scale suicidal ideation item; SCID = Structured Clinical Interview for *Diagnostic and Statistical Manual of Mental Disorders* suicide questions; SIQ = Suicidal Ideation Questionnaire; SIS = Beck's Suicide Intent Scale; SITBI = Self-Injurous Thoughts and Behaviours Interview; SSS = Suicide Score Scale; YSR = Youth Self-Report suicide item; SIQ-JR = Suicidal Ideation Questionnaire–Junior; ENIGMA = Enhancing NeuroImaging Genetics through Meta-Analysis. [*]For multi-item instruments, dimensionality reduction was carried out by extracting common factor scores.

HAM-D suicidal ideation items; $r = 0.35$ and 0.64, respectively; see Supplemental Table S1). In addition, test–retest reliability was lower for more ambiguous constructs such as suicidal gestures ($\kappa = 0.25$) in the SITBI. In general, interrater reliability for clinician-administered measures was strong for specialized multi-item instruments, such as the C-SSRS, SIS, SSI, and SITBI, and for clinical interviews, such as the KSADS, but moderate for single items on recent suicidal ideation from depression severity rating scales (such as the CDRS). Finally, internal consistency was strong for multi-item specialized instruments that focus on one construct (such as recent or lifetime ideation) but varied more across studies for the C-SSRS intensity of ideation subscale (see Supplemental Table S1).

The lowest concurrent validity measure identified was between the SIS and C-SSRS scale. Overall, mostly moderate-to-high correlation or concurrent validity scores (κ range: 0.15–0.97; r range: 0.21–0.94) between instruments were identified (Figure 1a). When examining the concurrent validity between self-report measures, the correlation ranged between 0.41 and 0.69, and the κ varied between 0.15 and 0.45, while for clinician-administered measures, the correlation varied between 0.21 and 0.94, and the κ varied between 0.42 and 0.97 (Figure 1a).

ENIGMA Meta-Analysis

Sample Description and Dimensionality Reduction

The average age across cohorts was 39 years ($SD = 16.3$). Cohorts had on average 40% male participants. The most commonly

available instruments were the suicidal ideation items from the MADRS, HAM-D, and BDI questionnaires. Multi-item items, such as the C-SSRS, QIDS, SSI, and SCID, were available for 32.4% of the participants. The majority of instruments were administered by a clinician or trained interviewer, but some self-reported measures were used (Table 2). For multi-item instruments, dimensionality reduction was carried out by extracting common factor scores using factor analysis (see Method section). Fit statistics of these models for each multi-item instrument within each cohort are summarized in Supplemental Table S4. These multi-item instruments typically measure more than one suicidal construct. For example, the C-SSRS and SSI include sections on protective factors. Thus, a single common factor might not represent the best model underlying the latent structure of these instruments, but it serves our purpose of

Table 2

Instruments Used to Assess Suicidal Thoughts and/or Behavior by the Different Cohorts: Self-Report and Clinician-/Interviewer-Administered Measures Are Presented Separately

Cohort name	Self-report	Clinician/interviewer administered
AFFDIS	BDI-II suicidal ideation item	MADRS and HAM-D suicidal ideation item
Chiba University	BDI-I suicidal ideation item, BDI-II suicidal ideation item	MINI suicide module
Duke/Durham VA	BDI suicidal ideation item, Beck Scale for Suicidal Ideation	
EPISCA (Leiden adolescents)	CDI suicidal ideation item, RCADS suicidal ideation item, YSR suicidal ideation item	
ETPB-STB	BDI suicidal ideation item	HAM-D suicidal ideation item, MADRS suicidal ideation item, Beck Scale for Suicidal Ideation
FIDMAG		HAM-D suicidal ideation item, MADRS suicidal ideation item
FOR 2107 Muenster	BDI suicidal ideation item	HAM-D suicidal ideation item
FOR 2107 Marburg	BDI suicidal ideation item	HAM-D suicidal ideation item
Grady Trauma Project Emory University		BDI suicidal ideation item, Columbia Suicide Severity Rating Scale
The University of Melbourne	QIDS suicidal ideation item, Suicidal Ideation Questionnaire	Columbia Suicide Severity Rating Scale, MADRS suicidal ideation item
University of Minnesota Adolescent MDD	IDAS, BDI suicidal ideation item	Children's Depression Rating Scale, KSADS suicide items
CHU Montpellier BICS study	BDI suicidal ideation item	IDS suicidal ideation item, Columbia Suicide Severity Rating Scale
CHU Montpellier IMPACT study	QIDS suicidal ideation item	IDS suicidal ideation item, Columbia Suicide Severity Rating Scale
CHU Montpellier Servier Study	BDI suicidal ideation item	HAM-D suicidal ideation item, Beck Scale for Suicidal Ideation
McGill University	BDI suicidal ideation item, QIDS suicidal ideation item, Scale for Suicidal Ideation	MADRS suicidal ideation item, HAM-D suicidal ideation item, Suicide Intent Scale, Columbia Suicide Severity Rating Scale
Moral Dilemma	QIDS suicidal ideation item	MADRS suicidal ideation item
Muenster Neuroimaging Cohort	BDI suicidal ideation item	HAM-D suicidal ideation item, MADRS suicidal ideation item
Fondazione Santa Lucia		Suicide Score Scale, Mini International Neuropsychiatric Interview suicidality module
San Raffaele Hospital	BDI suicidal ideation item, Beck Scale for Suicidal Ideation	HAM-D suicidal ideation item
SoCAT	BDI suicidal ideation item	HAM-D suicidal ideation item
South Africa		MADRS suicidal ideation item, DIGS suicide items
Stanford University adolescent MDD TIGER		C-SSRS, SITBI
Stanford University AGG/FAA		SCID suicide items, HAM-D suicidal ideation item
STRADL	QIDS suicidal ideation item	SCID suicide items
Sydney Bipolar Risk Study		MADRS suicide item, DIGS suicide items, KSADS suicide items
UCSF Adolescent MDD	BDI-II suicidal ideation item	Columbia Suicide Severity Rating Scale
Yale School of Medicine		Beck Scale for Suicidal Ideation, Suicide Intent Scale, Columbia Suicide Severity Rating Scale, Hamilton Depression Rating Scale, Children's Depression Rating Scale

Note. VA = Veteran Affairs; CDI = Children's Depression Inventory; STB = Suicidal Thoughts and Behaviour; MDD = Major Depressive Disorder; CHU = Centre Hospitalier Universitaire; UCSF = University of California, San Francisco; BDI = Beck Depression Inventory; MADRS = Montgomery–Asberg Depression Rating Scale; HAM-D = Hamilton Depression Rating Scale; MINI = Mini International Neuropsychiatric Interview; RCADS = Revised Children's Anxiety and Depression Scale; YSR = Youth Self-Report; QIDS = Quick Inventory of Depressive Symptomatology; IDAS-II = Inventory of Depression and Anxiety Symptoms; KSADS = Kiddie Schedule for Affective Disorders and Schizophrenia; IDS = Inventory of Depressive Symptomatology; DIGS = Diagnostic Interview for Genetics Studies; C-SSRS = Columbia Suicide Severity Rating Scale; SITBI = Self-Injurious Thoughts and Behaviors Interview; SCID = Structured Clinical Interview for *Diagnostic and Statistical Manual of Mental* Disorders.

dimensionality reduction while capturing the main underlying latent factor related to suicidality which these instruments assess.

Correlation Analyses

Results of our correlation analyses are summarized in Figure 1b. Full results are provided in Supplemental Table S5. As explained in the Method section, multi-item instruments were summarized using factor analysis for dimensionality reduction. From now on, when referring to the multi-item instruments listed in the Method section, we are referring to the common factor score obtained by the dimensionality reduction approach. Overall, moderate-to-high correlations (median $r = 0.44$) were observed among all the studied instruments (including single item and common factor of multi-item instruments). Nonetheless, the common factor of the SIS showed poor correlations (median $r \sim 0.20$) with most of the instruments tested. This result is not unexpected; the SIS was applied by a single cohort ($N = 16$) and assesses suicide intent during a suicide attempt and no suicidal ideation or behavior as the other instruments do (see Discussion section).

The instrument with the highest consistency (i.e., highest median-weighted correlations with other instruments) was the IDS-Clinician rated measure (median $r = 0.76$). However, few pairs of cohorts had data for this instrument. The C-SSRS and SSI instrument showed a very high concordance ($r = 0.83$, $N = 191$) with each other. In addition, there was a strong correlation between the HAM-D suicidal ideation item and the same item in the version of this questionnaire for children, the CDRS, but this was supported by a single cohort ($r = 0.94$, $N = 20$). The MADRS suicidal ideation item showed a high correlation with the HAM-D ($r = 0.67$, $N = 1,087$) and BDI ($r = 0.74$, $N = 844$) suicidal ideation items and with the SSI instrument common factor ($r = 0.67$, $N = 119$). The HAM-D and BDI suicide items showed only a moderate correlation ($r = 0.41$, $N = 2,555$) between them. Both of these measures were moderately correlated with the SSI ($r = 0.38$, $N = 429$ and $r = 0.36$, $N = 350$, respectively). Moderate-to-low correlations were observed for the group comprising child scales (YSR, RCADS, and Children's Depression Inventory), but these were only collected by one cohort. The MINI and SSS common factors showed a low correlation ($r = 0.12$), which was also supported by a single cohort only ($N = 64$). A secondary analysis excluding suicide deterrents or protective factors before dimensionality reduction showed practically the same results (Supplemental Figure 1).

Recent Versus Lifetime STB

Cohorts applied different instruments assessing STB with different time frames. For example, the C-SSRS could be used to assess lifetime, time since last assessment, and recent (past 2 weeks) suicidal behavior information, whereas other instruments might be worded around the past month, 2 weeks, week or even at the time of assessment. This is a potential source of heterogeneity for studies wishing to compare these measures. For this reason, we repeated the analyses only focusing on measures applied to a *recent* (within the past month) time frame and compared them to the results shown above. These analyses showed similar correlations overall. Notably, concordance between the C-SSRS and the HAM-D, as well as C-SSRS and BDI, was higher when focusing only on recent instruments (Figure 2a–b).

Comparing Clinician and Self-Report Scales

We gathered information on whether the distinct instruments were administered by a clinical interview or by self-report (Table 2), and the majority of instruments were administered by a clinician or trained interviewer. Within interviewer-administered scales, high correlations ($r > 0.7$) were observed between the SSI, C-SSRS, MADRS, and QIDS instruments. A similar result was observed between HAM-D and MADRS ($r = 0.67$). The SIS still showed a lower consistency with most other instruments (Figure 2c). For the self-report-based instruments, less data were available. Among the self-reported instruments, the BDI and IDAS, as well as the SSI and QIDS instruments showed a high concordance (Figure 2d). For interviewer-based measures, we were able to replicate the pattern from the main analyses: high correlations between single-item measures and measures assessing recent suicidal ideation (HAM-D, MADRS), strong correlations between detailed measures of STB (C-SSRS and SSI), and low correlations between the SIS and other measures. Nonetheless, we identified pairs of instruments such as the HAM-D and BDI whose low correlation in the main analysis might be explained by differences in administration (i.e., self-report vs. interviewer).

Construct-Specific Analyses

Overall, given the small number of cohorts with multi-item instruments, there were few instances where a comparison between construct factors was possible. Results were consistent with the full data analyses, for example, the SSI and C-SSRS suicide ideation constructs showed a high correlation ($r = 0.84$; Supplemental Figure 2). The MADRS single item showed a high correlation with the KSADS suicide attempt construct, but not with the C-SSRS suicide attempt construct. Single-item instruments, on average, showed a higher correlation with the ideation constructs of multi-item instruments compared to the other constructs. Consistent with our previous results, single item showed a very poor correlation with suicide intent constructs and moderate-to-low correlations with suicide attempt constructs (Supplemental Figure 2).

Discussion

Our study is a comprehensive assessment of how well different suicide risk assessment instruments relate to each other and the extent to which they can be used interchangeably. Harmonization reduces heterogeneity and increases power for discovery analyses, but also enables the assessment of the generalizability of studies and opens up the opportunity to investigate other aspects such as interactions and individual variation analysis (van Harmelen et al., 2021). Identification of correlates of suicide risk may be improved by increasing sample sizes, and by pooling data across studies to detect small effect sizes, which may result from large variance in underlying mechanisms. Our study aimed to examine the concurrent validity of instruments commonly used to assess STBs. To this end, we compared individual-level responses across questionnaires by pooling data from 27 cohorts belonging to the ENIGMA-MDD and ENIGMA-STB working groups. We compared our results to a systematic literature search across 225 studies.

A potential source of variance is the heterogeneity introduced by using different suicide risk assessment instruments that each

Figure 2
Sensitivity Results
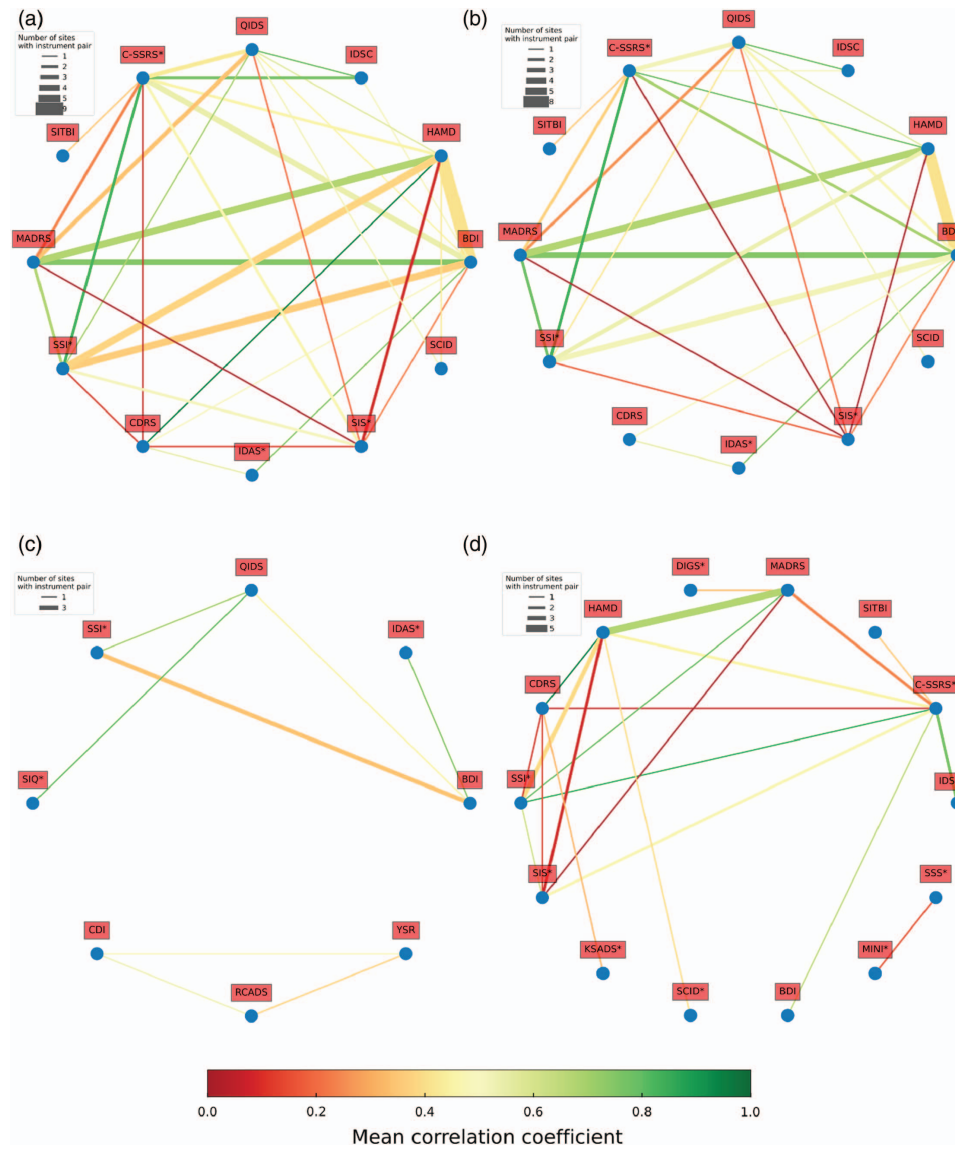
Note. Undirected acyclic graph shows the results for the meta-analysis of correlations of suicide risk assessment instruments across ENIGMA cohorts for the complete results (a); using only measures assessing recent (up to past month) suicidal behavior (b); instruments administered via self-report (c); or using interviewer administered instruments (d). Each node represents one of the instruments included in the study. Each edge color represents the sample size–weighted average correlation coefficient between two instruments. The thickness of the edge increases with the number of cohorts contributing to estimate the correlation. Generally speaking, the thicker the edge the more confidence in the correlation estimate. *For multi-item instruments, dimensionality reduction was carried out by extracting common factor scores. QIDS = Quick Inventory of Depressive Symptomatology suicidal ideation item; SITBI = Self-Injurious Thoughts and Behaviors Interview; MADRS = Montgomery–Asberg Depression Rating Scale suicidal ideation item; SSI = Scale for Suicidal Ideation; C-SSRS = Columbia Suicide Severity Rating Scale; CDRS = Children's Depression Rating Scale suicidal ideation item; IDAS = Inventory of Depression and Anxiety Symptoms; SIS = Beck's Suicide Intent Scale; SCID = Structured Clinical Interview for *Diagnostic and Statistical Manual of Mental* Disorders suicide questions; BDI = Beck Depression Inventory suicidal ideation item; SIQ = Suicidal Ideation Questionnaire; CDI = Children's Depression Inventory; RCADS = Revised Children's Anxiety and Depression Scale suicidal ideation item; YSR = Youth Self-Report suicide item; DIGS = Diagnostic Interview for Genetics Studies items on suicidal ideation and behavior; KSADS = Kiddie Schedule for Affective Disorders and Schizophrenia; MINI = Mini International Neuropsychiatric Interview suicidality module; SSS = Suicide Score Scale; HAM-D = Hamilton Depression Rating Scale item on suicidal ideation; ENIGMA = Enhancing NeuroImaging Genetics through Meta-Analysis; IDS = Inventory of Depressive Symptomatology.

measure is slightly different underlying phenomena. Both the results of our analysis and our literature search identified moderate-to-high correlations between the most commonly used instruments to assess suicidal ideation, including the BDI, SSI, HAM-D, and MADRS, and between multi-item instruments (such as the C-SSRS and SSI). These findings are consistent with another study which showed strong correlations between the SSI, BDI, and HAM-D (Desseilles et al., 2012). Nonetheless, our results were consistently more conservative than the literature (e.g., showing a lower degree of correlation). This could be explained in several ways. This might be evidence for publication bias whereby only positive and expected associations are published. Heterogeneity arising from the way these instruments are administered (i.e., time frames for suicidal behaviors and self-report vs. interview-based) might also affect the results. We performed sensitivity analyses testing for these factors. For example, the low correlations between BDI and HAM-D may be explained by the fact that the BDI is a self-report measure, whereas the HAM-D is typically administered via a clinician interview. There are studies showing that participants are more likely to disclose suicidal ideation in self-reported measures compared to clinical interviews (Kaplan et al., 1994; Yigletu et al., 2004). It is unclear whether a similar phenomenon underlies our results, given that the discrepancy is seen across cohorts rather than in individuals. Another example includes the correlation between the BDI and C-SSRS common factor, which became higher when focusing on the interview-based cohort only. Nonetheless, this explanation does not always hold true, for example, the MADRS (also interview-based) correlated well with the BDI.

Our unique methodology was chosen to integrate slightly different versions of instruments (i.e., self-reported vs. interview-administered). By extracting common factor scores for the multi-item instruments, we were able to obtain a measure of the underlying suicidal risk being measured by the instruments while reducing the need to adjust for slightly different wording or items between versions. This approach might be more conservative than those used in previous studies as it is focused on measuring suicide risk rather than a specific construct such as attempt or ideation. As such, instruments that mainly focused on a specific aspect of STB such as suicidal ideation (e.g., the SIQ) are expected to show a lower correlation with more broad instruments that assess a range of relevant behaviors, such as the C-SSRS.

It is worth noting that suicidal constructs may share a common liability, but they may also have partially independent etiologies. The SIS was the only measure with overall low measures of consistency in both our analyses and the literature search. Our findings do not necessarily imply that the SIS (or any of the measures studied) lacks utility, but rather that there is a lack of concordance with other measures which could be due to assessment of different constructs or other sources of heterogeneity. As such, future studies specifically focused on harmonization across suicidal measures should account for the design differences present in the SIS or single-item instruments and assess whether their inclusion provides valuable information. However, caution is warranted in interpreting the SIS findings as they were based on two cohorts only. Although our literature search identified a low concordance between the SIS and both the SIQ and the C-SSRS, a high concordance was reported between SIS and the K-SADS. These results are likely explained by the fact that the SIS is a questionnaire mainly focused on assessing intent of a past suicide attempt. This is further complicated by the intricate relationship between suicidal ideation, attempt, and actual suicide intent. Participants might engage in a suicide attempt with a relatively low intent to die. In fact, previous studies have identified that combining the SIS with other scales increases sensitivity and specificity for predicting suicide deaths (Stefansson et al., 2015). In a secondary analysis focusing on the *suicide intent* construct of multi-item instruments, we observed a very poor correlation between all single-item instruments and the suicide intent factors.

Our study represents a comprehensive approach to assess the concordance and reliability of commonly used suicide risk assessment instruments. Nonetheless, some limitations need to be considered when interpreting our findings. Our literature search was as exhaustive and systematic as was practicable; however, we cannot rule out the possibility that some relevant studies were excluded because they are not indexed in the databases we searched. Language and cultural differences between cohorts might also affect whether two instruments are concordant. Our study comprised predominantly English-speaking participants, but some cohorts included French-, Dutch-, and German-speaking participants. Limited research is available on whether language affects reporting of psychiatric symptomatology (Erkoreka et al., 2020). Participants might have undergone evaluation at different points in time in relation to the timing of suicidal behaviors or thoughts, and recall bias could lower the concordance between instruments. In fact, our sensitivity analyses showed that the time frame and administration of the scale used can affect how different measures compare with each other and that most single-item instruments seem to correlate better with suicide ideation constructs rather than attempt or intent constructs. When focusing on multi-item instruments, we performed dimensionality reduction using factor scores derived from FIML. Our approach was based on the fact that some instruments have multiple versions. Thus, our approach is not an exploratory or confirmatory factor analysis of the multi-item instruments used here. Performing such a study is outside the scope of this article as it would require complete harmonization of the questionnaires across cohorts and to focus solely on the multi-item instruments. Finally, we did not compare how single-item instruments on suicidality relate to broad depression or psychiatric interviews, future studies could tackle questions of convergence with broader constructs, such as depression, and of temporal stability over time, using a similar strategy as the one we employed here.

While our sensitivity analysis only distinguished between lifetime and recent time frames, it is possible that these effects exist even within recent time frames such as instruments assessing current versus past 2 week behaviors. These limitations might explain the lower concordance identified by our analysis compared to the literature synthesis, which studied a compendium of smaller albeit less heterogeneous studies. Additionally, while this was a large study, many sites had a distinct combination of measures collected (max $N \sim 2,500$); therefore, we did not have the power to perform additional sensitivity analyses in adults or adolescents only. Future studies should focus on addressing this, as there are clear factors associated with suicidality (e.g., mood reactivity) that are more prevalent during adolescence (Armey et al., 2015). We were also unable to stratify our analyses by the type of psychiatric diagnosis, while prior work found that the correlation between self-report and clinician-reported suicidal thoughts may differ across disorders (Gao et al., 2015; Kaplan et al., 1994). In addition, most sites in

the analysis included single items from depression severity rating scales, and detailed information from multi-item instruments was only available for one third of the participants. Finally, in this study, we have only included cross-sectional data, which means we are unable to shed light on how different measures assess changes in ideation or suicidal behavior over time. A recent study for instance has shown that the suicide item of the Ham-D was more strongly associated with suicidal ideation at follow-up, compared to the SSI and C-SSRS (McCall et al., 2021). Existing longitudinal studies suggest that suicidal ideation and suicidal behavior, measured here, are significant, weak predictors of death by suicide. Thus, this work may not shed light on the best tools to identify high-risk individuals and states and, instead, is focused on convergence of existing tools.

Overall, our results suggest that the most commonly used instruments show a moderate-to-high concordance. Use of different measures of suicidality might increase heterogeneity depending on the distinct dimensions and constructs assessed by each instrument. For research studies pooling data from multiple studies and interested in a broad suicidal risk construct, a strategy similar to ours could enable the implementation of composite scores by weighing more concordant measures heavily and penalizing less concordant measures. This would enable the inclusion of data from cohorts which have only weakly related instruments to increase power, but at the same time prevent bias by weighing these measures accordingly. In the absence of such an approach, large-scale research collaborations could focus on strictly defined suicide constructs such as suicidal ideation, attempt, and intent that are preferentially defined using the most common instruments such as—among others—the HAM-D, MADRS, SSI, C-SSRS. In the absence of a common instrument, or when using an instrument that we found to have low concordance, sensitivity analyses could be performed to assess whether significant results are robust, or at least consistent, after excluding cohorts using the least common instruments. We strongly recommend that future studies that plan to collect data on STBs include one or more of the instruments that showed strong correlations with instruments, such as the MADRS, SSI, and C-SSRS.

The moderate-to-high concordance between measures suggests that it is possible to harmonize across instruments and pool data from different research studies. However, it is important to acknowledge that the choice for a specific measure in a clinical setting depends on the construct that is of interest, as various instruments assess distinct constructs, for instance, intent, ideation, or attempt. Therefore, these measures cannot be blindly used interchangeably. These instruments may additionally differ in terms of detail and administration time, which may also impact which measures could be preferred in a clinical setting. For research settings, our results suggest that all of these factors will likely increase heterogeneity, and statistical adjustment or sensitivity analyses, along with increasing sample sizes, will be necessary for gaining insights into the etiology of suicide.

References

Achenbach, T., Achenbach, T. M., Achenbach, T. M., & Achenbach, T. (1991), *Manual for the youth self-report and 1991 profile*. https://www.scienceopen.com/document?vid=2d9fe2e3-04c5-4983-a979-2ed002bf63b7

Alicandro, G., Malvezzi, M., Gallus, S., La Vecchia, C., Negri, E., & Bertuccio, P. (2019). Worldwide trends in suicide mortality from 1990 to 2015 with a focus on the global recession time frame. *International Journal of Public Health*, *64*(5), 785–795. https://doi.org/10.1007/s00038-019-01219-y

Armey, M. F., Schatten, H. T., Haradhvala, N., & Miller, I. W. (2015). Ecological momentary assessment (EMA) of depression-related phenomena. *Current Opinion in Psychology*, *4*, 21–25. https://doi.org/10.1016/j.copsyc.2015.01.002

Bates, T. C. (2018). *Umx: A helper package for structural equation modeling in OpenMx* [Doctoral dissertation]. University of Edinburgh.

Bates, T. C., Neale, M. C., & Maes, H. H. (2016). Umx: A library for structural equation and twin modelling in R. *Journal of Statistical Software*, *22*(1), 27–41. https://doi.org/10.1017/thg.2019.2

Beck, A., Schuyler, D., & Herman, I. (1974). *Development of suicidal intent scales*. Charles Press Publishers.

Beck, A. T., Kovacs, M., & Weissman, A. (1979). Assessment of suicidal intention: The scale for suicide ideation. *Journal of Consulting and Clinical Psychology*, *47*(2), 343–352. https://doi.org/10.1037/0022-006X.47.2.343

Beck, A. T., Steer, R. A., & Ranieri, W. F. (1988). Scale for suicide ideation: Psychometric properties of a self-report version. *Journal of Clinical Psychology*, *44*, 499–505. https://doi.org/10.1002/1097-4679(198807)44:4<499::AID-JCLP2270440404>3.0.CO;2-6

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An Inventory for Measuring Depression. *Archives of General Psychiatry*, *4*, 561–571. https://doi.org/10.1001/archpsyc.1961.01710120031004

Campos, A. I., Thompson, P. M., Veltman, D. J., Pozzi, E., van Veltzen, L. S., Jahanshad, N., Adams, M. J., Baune, B. T., Berger, K., Brosch, K., Bülow, R., Connolly, C. G., Dannlowski, U., Davey, C. G., de Zubicaray, G. I., Dima, D., Erwin-Grabner, T., Evans, J. W., Fu, C. H. Y., … ENIGMA-MDD Working Group. (2020). Brain correlates of suicide attempt in 18,925 participants across 18 international cohorts. *Biological Psychiatry*, *90*(4), 243–252. https://doi.org/10.1016/j.biopsych.2021.03.015

Chorpita, B. F., Yim, L., Moffitt, C., Umemoto, L. A., & Francis, S. E. (2000). Assessment of symptoms of DSM-IV anxiety and depression in children: A revised child anxiety and depression scale. *Behaviour Research and Therapy*, *38*(8), 835–855. https://doi.org/10.1016/S0005-7967(99)00130-8

Desseilles, M., Perroud, N., Guillaume, S., Jaussent, I., Genty, C., Malafosse, A., & Courtet, P. (2012). Is it valid to measure suicidal ideation by depression rating scales? *Journal of Affective Disorders*, *136*(3), 398–404. https://doi.org/10.1016/j.jad.2011.11.013

Erkoreka, L., Ozamiz-Etxebarria, N., Ruiz, O., & Ballesteros, J. (2020). Assessment of psychiatric symptomatology in bilingual psychotic patients: A systematic review and meta-analysis. *International Journal of Environmental Research and Public Health*, *17*(11), Article 4137. https://doi.org/10.3390/ijerph17114137

First, M. B. (1997). *Structured clinical interview for DSM-IV axis I disorders SCID-I: Clinician version, scoresheet*. American Psychiatric Press.

Gao, K., Wu, R., Wang, Z., Ren, M., Kemp, D. E., Chan, P. K., Conroy, C. M., Serrano, M. B., Ganocy, S. J., & Calabrese, J. R. (2015). Disagreement between self-reported and clinician-ascertained suicidal ideation and its correlation with depression and anxiety severity in patients with major depressive disorder or bipolar disorder. *Journal of Psychiatric Research*, *60*, 117–124. https://doi.org/10.1016/j.jpsychires.2014.09.011

Hagberg, A., Swart, P., & S Chult, D. (2008). *Exploring network structure, dynamics, and function using NetworkX* (No. LA-UR-08-05495; LA-UR-08-5495). Los Alamos National Laboratory.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, *23*(1), 56–62. https://doi.org/10.1136/jnnp.23.1.56

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., Del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Innamorati, M., Pompili, M., Lester, D., Tatarelli, R., & Girardi, P. (2008). Recreational drug use and suicidality among Italian young adults. *Journal of Addictive Diseases*, *27*(4), 51–59. https://doi.org/10.1080/10550880 802324796

Kaplan, M. L., Asnis, G. M., Sanderson, W. C., Keswani, L., De Lecuona, J. M., & Joseph, S. (1994). Suicide assessment: Clinical interview vs. self-report. *Journal of Clinical Psychology*, *50*(2), 294–298. https://doi.org/10 .1002/1097-4679(199403)50:2<294::AID-JCLP2270500224>3.0.CO;2-R

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., Williamson, D., & Ryan, N. (1997). Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): Initial reliability and validity data. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(7), 980–988. https:// doi.org/10.1097/00004583-199707000-00021

McCall, W. V., Porter, B., Pate, A. R., Bolstad, C. J., Drapeau, C. W., Krystal, A. D., Benca, R. M., Rumble, M. E., & Nadorff, M. R. (2021). Examining suicide assessment measures for research use: Using item response theory to optimize psychometric assessment for research on suicidal ideation in major depressive disorder. *Suicide & Life-Threatening Behavior*, *51*(6), 1086–1094. https://doi.org/10.1111/sltb.12791

Montgomery, S. A., & Asberg, M. (1979). A new depression scale designed to be sensitive to change. *The British Journal of Psychiatry*, *134*(4), 382–389. https://doi.org/10.1192/bjp.134.4.382

Mullins, N., Kang, J., Campos, A. I., Coleman, J. R. I., Edwards, A. C., Galfalvy, H., Levey, D. F., Lori, A., Shabalin, A., Starnawska, A., Su, M. H., Watson, H. J., Adams, M., Awasthi, S., Gandal, M., Hafferty, J. D., Hishimoto, A., Kim, M., Okazaki, S., … the VA Million Veteran Program. (2022). Dissecting the shared genetic architecture of suicide attempt, psychiatric disorders, and known risk factors. *Biological Psychiatry*, *91*(3), 313–327. https://doi.org/10.1016/j.biopsych.2021.05.029

Nock, M. K., Holmberg, E. B., Photos, V. I., & Michel, B. D. (2007). Self-injurious thoughts and behaviors interview: Development, reliability, and validity in an adolescent sample. *Psychological Assessment*, *19*(3), 309–317. https://doi.org/10.1037/1040-3590.19.3.309

Nurnberger, J. I., Jr., Blehar, M. C., Kaufmann, C. A., York-Cooler, C., Simpson, S. G., Harkavy-Friedman, J., Severe, J. B., Malaspina, D., Reich, T., & the NIMH Genetics Initiative. (1994). Diagnostic interview for genetic studies. Rationale, unique features, and training. *Archives of General Psychiatry*, *51*(11), 849–859. https://doi.org/10.1001/archpsyc .1994.03950110009002

Posner, K., Brown, G. K., Stanley, B., Brent, D. A., Yershova, K. V., Oquendo, M. A., Currier, G. W., Melvin, G. A., Greenhill, L., Shen, S., & Mann, J. J. (2011). The Columbia–Suicide severity rating scale: Initial validity and internal consistency findings from three multisite studies with adolescents and adults. *American Journal of Psychiatry*, *168*, 1266–1277. https://doi.org/10.1176/appi.ajp.2011.10111704

Poznanski, E. O., & Mokros, H. B. (1996). *Children's depression rating scale, revised (CDRS-R)*. Western Psychological Services.

Reynolds, W. M. (1987). *Suicidal ideation questionnaire (SIQ)*. Psychological Assessment Resources. http://www.v-psyche.com/doc/Clinical%20Test/ Suicidal%20Ideation%20Questionnaire.doc

Rush, A., Giles, D. E., Schlesser, M. A., Fulton, C. L., Weissenburger, J., & Burns, C. (1986). The inventory for depressive symptomatology (IDS): Preliminary findings. *Psychiatry Research*, *18*(1), 65–87. https://doi.org/ 10.1016/0165-1781(86)90060-0

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. C. (1998). The mini-international neuropsychiatric interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry*, *59*(20, Suppl. 20), 22–33.

Stefansson, J., Nordström, P., Runeson, B., Åsberg, M., & Jokinen, J. (2015). Combining the suicide intent scale and the karolinska interpersonal violence scale in suicide risk assessments. *BMC Psychiatry*, *15*(1), Article 226. https://doi.org/10.1186/s12888-015-0607-6

van Harmelen, A.-L., Schmaal, L., & Blumberg, H. P. (2021). Imaging suicidal thoughts and behavior: The promise of computational models. *Neuropsychopharmacology*, *46*(1), 219–220. https://doi.org/10.1038/ s41386-020-00841-2

van Velzen, L. S., Dauvermann, M. R., Colic, L., Villa, L. M., Savage, H. S., Toenders, Y. J., Zhu, A. H., Bright, J. K., Campos, A. I., Salminen, L., Ambrogi, S., Ayesa-Arriola, R., Banaj, N., Başgöze, Z., Bauer, J., Blair, K., Blair, R. J., Brosch, K., Cheng, Y., … Schmaal, L. (2022). Structural brain alterations associated with suicidal thoughts and behaviors in young people: Results from 21 international studies from the ENIGMA Suicidal Thoughts and Behaviours consortium. *Molecular Psychiatry*, *27*(11), 4550–4560. https://doi.org/10.1038/s41380-022-01734-0

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … the SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Watson, D., O'Hara, M. W., Naragon-Gainey, K., Koffel, E., Chmielewski, M., Kotov, R., Stasik, S. M., & Ruggero, C. J. (2012). Development and validation of new anxiety and bipolar symptom scales for an expanded version of the IDAS (the IDAS-II). *Assessment*, *19*(4), 399–420. https:// doi.org/10.1177/1073191112449857

World Health Organization. (1997). *CIDI-auto Version 2.1: Administrator's guide and reference*.

World Health Organization. (2014). *Preventing suicide: A global imperative*.

Yigletu, H., Tucker, S., Harris, M., & Hatlevig, J. (2004). Assessing suicide ideation: Comparing self-report versus clinician report. *Journal of the American Psychiatric Nurses Association*, *10*(1), 9–15. https://doi.org/10 .1177/1078390303262655