# The reliability and clinical utility of ICD-11 schizoaffective disorder: A field trial

Destiny L. Peterson [a], Christopher A. Webb [a], Jared W. Keeley [b,*], Wolfgang Gaebel [c], Juergen Zielasek [d], Tahilia J. Rebello [e,f], Rebeca Robles [g], Chihiro Matsumoto [h], Cary S. Kogan [i], Maya Kulygina [j], Saeed Farooq [k], Michael F. Green [l], Peter Falkai [m], Alkomiet Hasan [m], Silvana Galderisi [n], Veronica Larach [o], Valery Krasnov [j], Geoffrey M. Reed [e,p]

[a] Mississippi State University, Starkville, MS, USA
[b] Virginia Commonwealth University, Richmond, VA, USA
[c] Department of Psychiatry and Psychotherapy, Medical Faculty, LVR-Klinikum, Heinrich-Heine University, Germany
[d] LVR-Institute for Healthcare Research, Cologne, Medical Faculty, Heinrich-Heine University, Düsseldorf, Germany
[e] Department of Psychiatry, Columbia University College of Physicians and Surgeons, New York, NY, USA
[f] New York State Psychiatric Institute, New York, NY, USA
[g] National Institute of Psychiatry Ramón de la Fuente Muñiz, Mexico City, Mexico
[h] Japanese Society of Psychiatry and Neurology, Tokyo, Japan
[i] School of Psychology, University of Ottawa, Ottawa, Ontario, Canada
[j] Moscow Research Institute of Psychiatry, National Medical Research Centre for Psychiatry and Narcology, Moscow, Russian Federation
[k] Research Institute for Primary Care & Health Sciences, Keele University, Staffordshire, United Kingdom
[l] Department of Psychiatry and Biobehavioral Sciences, Geffen School of Medicine, University of California, Los Angeles, CA, USA
[m] Department of Psychiatry and Psychotherapy, Klinikum der Universität München, München, Germany
[n] University of Campania Luigi Vanvitelli, Naples, Italy
[o] Universidad Andres Bello, Santiago, Chile
[p] Department of Mental Health and Substance Abuse, World Health Organization, Geneva, Switzerland

## ARTICLE INFO

## ABSTRACT

A major goal for the revision of the *International Classification of Diseases and Related Health Problems, Tenth Edition* (ICD-10) is to increase the clinical utility of the diagnostic system. Schizoaffective disorder has a history of poor diagnostic reliability due to the similarities and overlap in symptoms that it shares with other disorders, especially primary psychotic and mood disorders. The present study was part of the case-controlled field trials for ICD-11 and examines how the proposed changes for schizoaffective disorder may improve differential diagnosis and diagnostic accuracy. Clinicians from around the globe ($n = 873$) were provided with either ICD-10 or ICD-11 diagnostic guidelines and asked to apply them to case vignettes comparing schizoaffective disorder to schizophrenia and mood disorders with psychotic symptoms. Participants were asked to respond to follow-up diagnostic questions to determine which components of the diagnostic guidelines affected diagnostic accuracy. Overall, clinicians showed small improvements in accurately diagnosing vignettes using ICD-11 over ICD-10. Results suggest the discrepancy in diagnosing schizoaffective disorder is related primarily to the presence of mood symptoms and discrepancies about whether those symptoms are more consistent with schizoaffective disorder or a mood disorder diagnosis. Continuing to identify ways to more accurately capture this symptom picture will be important in the future as well as systematic efforts to educate clinicians about differential diagnosis.

© 2019 Elsevier B.V. All rights reserved.

## 1. Introduction

The World Health Organization is nearing completion of the Eleventh Revision of the *International Classification of Diseases and Related Health Problems* (ICD-11), which is due to be approved by the World Health Assembly in 2019. A primary objective for changes to the ICD-11 classification of mental disorders as compared to ICD-10 has been to increase clinical utility (Reed, 2010). The World Health Organization (WHO) has defined clinical utility as the ability of diagnostic concepts to aid in communication and understanding about a diagnosis, their ability to be applied in clinical settings, and their ability to aid in prognostic knowledge and treatment planning (Reed, 2010). This definition

---

includes traditional concepts of diagnostic reliability and accuracy, which are necessary for meeting other functions like communication. WHO aimed to give practitioners a tool that would better align with their practical needs related to diagnosing and treating mental disorders (First, 2010; Reed, 2010). It is important for clinicians to have an understanding of the diagnostic concepts they use and for that understanding to be applicable across world regions.

The diagnostic guidelines for schizoaffective disorder are an area of particular interest in the present revision process (Gaebel, 2012). Schizoaffective disorder has longstanding issues with validity and reliability, in addition to continued questions regarding how the diagnosis should be most accurately classified (Malaspina et al., 2013; Malhi et al., 2008). A meta-analysis of 49 publications revealed that the mean test-retest reliability of schizoaffective disorder ($\kappa =$ 0.50, 95%CI 0.40–0.59) was consistently lower than its differential diagnoses of Schizophrenia ($\kappa = 0.69$, 95%CI 0.64–0.74), Bipolar Disorder ($\kappa = 0.77$, 95%CI 0.73–0.82) and Unipolar Depression ($\kappa = 0.73$, 95%CI 0.66–0.79; Santelmann et al., 2015). Schizoaffective disorder is characterized by psychotic symptoms occurring concomitantly with mood symptoms of either polarity. This combination of symptoms can be challenging for clinicians when differentiating this diagnosis from diagnostic categories characterized mainly by psychotic symptoms (as seen with Schizophrenia and Other Primary Psychotic Disorders) or from mood disorders with psychotic symptoms. One factor that potentially contributes to this confusion is that psychotic disorders appear to exist upon a continuum of symptom severity as opposed to being discrete diagnostic entities, and schizoaffective disorder adds an additional dimension of the presence of mood symptoms (Jonathan et al., 2013; Kendell and Jablensky, 2003; Malhi et al., 2008).

ICD-11 defines schizoaffective disorder as having all the features required for Schizophrenia in concurrence or within a few days of a mood episode. Psychotic and mood symptoms must be present for at least 4 weeks, and symptoms are not better accounted for by another condition (Gaebel et al., 2012). ICD-11 guidelines are different from the criteria in DSM-5, which conceptualize schizoaffective disorder as a longitudinal (principally life-long) disease course and not a cross-sectional observation at the time of clinical contact as in ICD-11. In addition, DSM-5 requires that hallucinations or delusions be present for at least two weeks *without* the presence of mood symptoms (APA, 2013). The WHO Working Group for Schizophrenia and Other Primary Psychotic Disorders considered the longitudinal approach problematic because in clinical practice it may be difficult for patients or clinicians to ascertain the presence of past psychotic or mood episodes and their temporal relationships with sufficient diagnostic accuracy.

In contrast, ICD-10 defines schizoaffective disorder as the presence of at least one of the core symptoms ("a–d" in ICD-10) of Schizophrenia and two concurrent symptoms of a mood episode (WHO, 1992). A consequence of this structure was that the episode of illness did not meet diagnostic requirements for either Schizophrenia or a Depressive or Manic Episode. Thus, the presence of symptoms not sufficient to diagnose Schizophrenia or mood disorders could be diagnosed as schizoaffective disorder in ICD-10. On the other hand, the ICD-10 stated that if both schizophrenic and affective symptoms developed together and were evenly balanced, the diagnosis of schizoaffective disorder (F25.-) should be made, even if the schizophrenic symptoms by themselves would have justified the diagnosis of Schizophrenia (WHO, 1992). The ICD-11 sought to resolve such ambiguities and inconsistencies by requiring that all the requirements of Schizophrenia and at least a Depressive Episode of moderate severity or a Manic or Mixed Episode must be present. Subthreshold symptoms can be specified using the new ICD-11 symptom qualifiers for primary psychotic disorders (Keeley and Gaebel, 2018), avoiding the need to classify the presence of subthreshold psychotic states in mood disorders or subthreshold mood states in psychotic disorders as schizoaffective disorder. In diagnoses of mood disorders with psychotic symptoms, the diagnosis can

shift to schizoaffective disorder if psychotic symptoms reach the threshold for a diagnosis of Schizophrenia.

The current study is part of a larger set of WHO studies investigating the clinical utility of various diagnostic areas (Keeley et al., 2016). This study focuses on schizoaffective disorder. Specifically, our research question was whether clinicians would be able to accurately diagnose schizoaffective disorder and differentiate it from other similar symptom profiles (i.e., Schizophrenia, Depressive Episode with Psychotic Symptoms, Manic Episode with Psychotic Symptoms) using case vignettes. Improved diagnostic accuracy (identifying the correct diagnosis) and reliability (coming to the same diagnostic conclusion) relative to ICD-10 would support the clinical utility of the ICD-11 guidelines.

## 2. Method

### 2.1. Participants

Participants were drawn from the Global Clinical Practice Network (GCPN; Reed et al., 2015), a worldwide network of mental health and allied professionals established for the purpose of the ICD-11 case-controlled field trials. Professionals were invited to join the GCPN through professional listservs; national and regional professional associations; international and national conferences in psychology, psychiatry, and related disciplines; and professional word-of-mouth. The study was administered in six languages: Chinese, English, French, Japanese, Russian, and Spanish. At the time the study was launched, 9323 GCPN members were eligible for the study and were invited to participate. To qualify, the participant must have endorsed a self-rated proficiency of being either advanced or fluent in one of the six languages of the study, and was seeing patients or engaged in clinical supervision. If a participant endorsed fluency in multiple languages, the participant was assigned to participate in his or her primary professional language, if available. Of the eligible and valid participants, 2629 (28.2%) responded to the survey link and began the study. Of those who agreed to participate, most (n = 2330; 88.6%) completed the study (25.0% of the total invited).

The present study examined the portion of the overall sample that completed parts of the study relevant to the diagnosis of schizoaffective disorder. Thus, the final sample for the current analyses consisted of 873 mental health and allied professionals representing 81 different nationalities. The regional distribution of participants can be found in Table 1, along with their gender, profession, and mean age and years of experience.

### 2.2. Materials

The survey was administered through Qualtrics, a web-based survey program. The materials in the survey included (a) diagnostic guidelines for ICD-11 and ICD-10 (psychotic disorders, depressive and manic episodes with psychotic features, and acute stress reaction), (b) a series of 10 vignettes for diagnostic comparisons (only 4 of which are presented here), and (c) a variety of diagnostic and clinical utility questions regarding the vignettes. Specifically, participants provided a diagnosis for a vignette and then answered questions about the presence or absence of each diagnostic guideline for the diagnosis they selected. The specific queries were adapted from the wording of ICD-10 or ICD-11 diagnostic guidelines. Results from the clinical utility questions are not included in this paper. For more information on study translation procedures, see Keeley et al. (2016).

Members of the WHO Working Group for Schizophrenia and Other Primary Psychotic Disorders developed the vignettes. Working Group members wrote vignettes based on actual patients (obscuring any personal details for confidentiality), which were then adapted by the corresponding author (JK) to meet the required experimental constraints (e.g., number and type of symptoms in the vignette). We explicitly developed the vignettes to reflect real cases to maximize applicability of

**Table 1**
Participant demographics by language.

| | Chinese $n = 100$ | English $n = 364$ | French $n = 78$ | Japanese $n = 118$ | Russian $n = 108$ | Spanish $n = 105$ |
|---|---|---|---|---|---|---|
| | $f$ (%) | $f$ (%) | $f$ (%) | $f$ (%) | $f$ (%) | $f$ (%) |
| **Region** | | | | | | |
| Africa | 0 | 13 (3.6) | 6 (7.7) | 0 | 0 | 0 |
| North America | 0 | 90 (24.7) | 4 (5.1) | 0 | 0 | 1 (1.0) |
| South America | 0 | 22 (6.0) | 1 (1.3) | 0 | 0 | 74 (70.5) |
| Eastern Mediterranean | 0 | 20 (5.5) | 8 (10.3) | 0 | 0 | 0 |
| Europe | 0 | 156 (42.9) | 59 (75.6) | 0 | 106 (98.1) | 30 (28.6) |
| South Eastern Asia | 0 | 42 (11.5) | 0 | 0 | 0 | 0 |
| Asian Western Pacific | 100 (100) | 0 | 0 | 117 (99.2) | 0 | 0 |
| Oceania Western Pacific | 0 | 20 (5.5) | 0 | 0 | 0 | 0 |
| Other | 0 | 1 (0.3) | 0 | 1 (0.8) | 2 (1.9) | 0 |
| **Gender** | | | | | | |
| Male | 63 (63.0) | 176 (48.4) | 46 (59.0) | 93 (78.8) | 65 (60.2) | 58 (55.2) |
| Female | 37 (37.0) | 188 (51.6) | 32 (41.0) | 25 (21.2) | 43 (39.8) | 46 (43.8) |
| **Profession** | | | | | | |
| Counseling | 4 (4.0) | 26 (7.1) | 0 | 3 (2.5) | 1 (0.9) | 0 |
| Psychiatry | 90 (90.0) | 156 (42.9) | 54 (69.2) | 96 (81.4) | 95 (88.0) | 40 (38.1) |
| Nursing | 0 | 11 (3.0) | 2 (2.6) | 1 (0.8) | 0 | 1 (1.0) |
| Psychology | 4 (4.0) | 134 (36.8) | 22 (28.2) | 14 (11.9) | 10 (9.3) | 51 (48.6) |
| Social work | 2 (2.0) | 5 (1.4) | 0 | 2 (1.7) | 0 | 2 (1.9) |
| Other | 0 | 32 (8.8) | 0 | 2 (1.7) | 2 (1.9) | 11 (10.2) |
| | $M$ ($SD$) | $M$ ($SD$) | $M$ ($SD$) | $M$ ($SD$) | $M$ ($SD$) | $M$ ($SD$) |
| Age | 40.63 (9.50) | 45.61 (11.15) | 46.01 (12.12) | 45.99 (11.01) | 42.90 (11.90) | 47.41 (11.85) |
| Years of Experience | 12.11 (8.91) | 14.29 (9.94) | 16.42 (10.69) | 14.65 (10.12) | 17.10 (11.25) | 18.62 (10.93) |

the findings to clinical practice, given the artificial nature of vignette studies (Evans et al., 2015). Vignettes are available from the corresponding author upon request.

Each vignette was developed to meet all the diagnostic requirements for a specific disorder. To determine if the intended characteristics were indeed present, a group of content experts pretested vignettes to confirm the presence or absence of required features. Pretesting also confirmed expert consensus about the correct diagnosis for the case. Based on the results of the pretest, some vignettes were modified to clarify content or ensure that specific symptoms were recognizable.

### 2.3. Procedure

Participants received an e-mail invitation to participate in the study through Qualtrics. Upon entry to the study, participants were randomly assigned to view and use either the ICD-10 or ICD-11 diagnostic guidelines. Once participants had reviewed the guidelines, they were randomly assigned to one of eight comparisons (Comparisons 1–3 are the focus of this paper). Each comparison reflected a specific diagnostic distinction between similar disorders. Within each comparison, participants viewed two vignettes counterbalanced for presentation order. Once the participants had viewed a vignette, they were asked to provide a diagnosis from a preset list based on the diagnostic guidelines they were assigned. Participants were given the option of writing in another diagnosis if their selection was not present, or indicating that no diagnosis was warranted. Participants then rated the specific presence or absence of each diagnostic guideline for the diagnosis they gave. At that point, clinicians were given the option of selecting a different final diagnosis. Participants then rated the diagnostic and clinical utility questions. Participants then completed the sequence again for a second vignette. The description of the specific comparisons included in this study follows.

#### 2.3.1. Comparison 1 – schizophrenia (Vignette 1) vs. schizoaffective disorder (Vignette 2)

The first comparison examined clinicians' ability to differentiate Schizophrenia from schizoaffective disorder. More specifically, Vignette 1 had sufficient psychotic symptoms for a diagnosis of Schizophrenia

but did not evidence any mood symptoms, whereas Vignette 2 included sufficient symptoms for a diagnosis of a Depressive Episode that overlapped in time with symptoms of Schizophrenia.

#### 2.3.2. Comparison 2 – schizoaffective disorder (Vignette 2) vs. depressive episode with psychotic symptoms (Vignette 3A)

The second comparison tested clinicians' ability to differentiate schizoaffective disorder from a Depressive Episode with Psychotic Symptoms. The difference between the two cases was that psychotic symptoms for Vignette 3A did not meet the full requirements for Schizophrenia.

#### 2.3.3. Comparison 3 – schizoaffective disorder (Vignette 2) vs. manic episode with psychotic symptoms (Vignette 3B)

Comparison 3 mirrored Comparison 2, but changed the mood episode in Vignette 3B from depressive to manic. Otherwise, the distinction remained the same.

## 3. Results

### 3.1. Diagnostic agreement

Diagnostic agreement for schizoaffective disorder was calculated for both ICD-11 and ICD-10 diagnostic guidelines. Whereas Cohen's kappa for ICD-11 schizoaffective disorder was 0.38 (0.32–0.45) it was 0.27 (0.21–0.34) for ICD-10. These values show improvement of ICD-11 over ICD-10, which is encouraging given low diagnostic agreement for schizoaffective disorder. For comparison, the kappas for Schizophrenia, Depressive Episode with Psychotic Symptoms, and Manic Episode with Psychotic Symptoms were 0.42 (0.34–0.49), 0.61 (0.53–0.68), and 0.87 (0.83–0.92), respectively, for ICD-11; for ICD-10 they were 0.34 (0.26–0.41), 0.66 (0.60–0.72), and 0.91 (0.88–0.95).

### 3.2. Comparison 1 – schizophrenia vs. schizoaffective disorder

Participants differentiated schizophrenia from schizoaffective disorder regardless of which diagnostic system they used, ICD-11 $\chi^2(2) = 66.35$, $p < .001$; ICD-10 $\chi^2(2) = 43.94$, $p < .001$. Despite the distinction, participants evidenced low diagnostic accuracy for Vignette 2. For ICD-

11, they diagnosed the first vignette relatively well as Schizophrenia (72.14% correct), but struggled to diagnose the second vignette as schizoaffective disorder (only 44.29% correct). However, participants using the ICD-10 performed significantly worse at the distinction relative to ICD-11, $G^2(4) = 11.50$, $p < .05$. The difference in performance was not attributable to Vignette 1, $\chi^2(2) = 2.11$, $ns$, but rather from differences on Vignette 2, $\chi^2(2) = 9.36$, $p < .01$. Diagnoses for Comparison 1 can be found in Table 2.

### 3.3. Comparison 2 – schizoaffective disorder vs. depressive episode with psychotic symptoms

For the second comparison, participants using both diagnostic systems differentiated schizoaffective disorder from a depressive episode with psychotic symptoms, ICD-11 $\chi^2(2) = 141.79$, $p < .001$; ICD-10 $\chi^2(2) = 206.64$, $p < .001$. As with Comparison 1, clinicians struggled with the diagnosis of Vignette 2. However, in this comparison the two systems were not statistically different in the diagnosis of Vignette 2 ($\chi^2(2) = 3.28$, $ns$), Vignette 3A ($\chi^2(2) = 5.19$, $ns$), or overall ($G^2(4) = 8.78$, $ns$). Diagnoses for Comparison 2 can be found in Table 3.

### 3.4. Comparison 3 – schizoaffective disorder vs. manic episode with psychotic symptoms

Participants differentiated schizoaffective disorder from a Manic Episode with Psychotic Symptoms on both diagnostic systems, ICD-11 $\chi^2(2) = 210.25$, $p < .001$; ICD-10 $\chi^2(2) = 224.01$, $p < .001$. Vignette 2 included depressive symptoms in the presentation of schizoaffective disorder, so the differentiation of vignettes may be based mostly on clinicians' recognition of the difference between depressive and manic features. That said, participants using ICD-11 were slightly more accurate diagnosing Vignette 2 than those using ICD-10, $\chi^2(2) = 6.31$, $p < .05$. There was no difference between systems on Vignette 3B, $\chi^2(2) = 3.47$, $ns$. Diagnoses for Comparison 3 can be found in Table 4.

### 3.5. Symptom endorsement for misdiagnoses

We also examined participants' endorsement of the diagnostic requirements when they gave an incorrect diagnosis for Vignette 2. We combined the follow-up diagnostic questions for the three comparisons in an effort to determine why clinicians provided an incorrect diagnosis. The most common misdiagnosis across comparisons was diagnosing Schizophrenia when schizoaffective disorder was the correct diagnosis. Across all three comparisons, 131 (31.95%) of the 410 participants using ICD-11 misdiagnosed Vignette 2 with Schizophrenia. Overall, there was high agreement on the requirement that at least two symptoms of Schizophrenia were present ($n = 127$, 96.95%), at least one symptom was a core symptom ($n = 130$, 99.24%), symptoms were present for at least one month ($n = 126$, 96.18%), and medical causes and substance use were ruled out ($n = 117$, 89.31%; $n = 115$, 87.79%). However, discrepancies appeared to be related to the presence of sufficient mood symptoms that would warrant another diagnosis. Eighteen (13.74%)

participants agreed that there were sufficient mood symptoms to warrant another diagnosis (but gave a diagnosis of Schizophrenia anyway), and 95 (72.52%) denied the presence of sufficient mood symptoms.

A slightly higher proportion of participants using ICD-10 misdiagnosed Vignette 2 as Schizophrenia (194 of 463; 41.90%). Of these, participants agreed on the presence of psychotic symptoms ($n = 191$, 98.45%), their duration being greater than a month ($n = 188$, 96.91%), and that the symptoms were not due to another medical condition ($n = 185$, 95.36%) or substance use ($n = 182$, 93.81%). However, there were discrepancies about the presence of mood symptoms, with 84 (43.30%) participants indicating that there were mood symptoms present, 90 (46.39%) denying the presence of mood symptoms, and 18 (9.28%) expressing uncertainty.

Although to a lesser degree than Schizophrenia, Vignette 2 was also misdiagnosed as a Depressive Episode with Psychotic Symptoms ($n = 72$, 17.56%) using ICD-11. There was general consensus on the presence of a sufficient number of depressive symptoms ($n = 69$, 95.83%), symptom duration ($n = 70$, 97.22%), symptoms from the affective cluster (i.e., depressed mood or loss of interest/pleasure, $n = 72$, 100%), presence of hallucinations or delusions ($n = 72$, 100%), functional impairment ($n = 70$, 97.22%), and medical causes and substance use ruled out ($n = 59$, 81.94%; $n = 65$, 90.28%). There appeared to be the most discrepancies related to whether the psychotic symptoms occurred outside of the mood episode with 52 (72.22%) participants endorsing that psychotic symptoms did not occur outside of the mood episode, 14 (19.44%) stating they were unsure, and 6 (8.33%) correctly endorsing that they did.

A misdiagnosis of a Depressive Episode with Psychotic Symptoms for participants using ICD-10 was comparable to those using ICD-11 ($n = 78$). However, there was less consensus on the presence of the required symptoms. Participants generally agreed that the core symptoms of a severe depressive episode were present ($n = 61$, 78.21%), along with additional depressive symptoms ($n = 59$, 75.64%), for the required duration ($n = 66$, 84.6%), that significantly impacted the person's functioning ($n = 59$, 75.64%), along with hallucinations or delusions ($n = 69$, 88.46%), not caused by another medical problem ($n = 61$, 78.21%) or substance use ($n = 65$, 83.33%). Once again, there was substantial disagreement about whether the symptoms met the requirements for a psychotic disorder, with 22 (28.21%) participants saying yes, 24 (30.77%) saying no, and 24 (30.77%) expressing uncertainty.

## 4. Discussion

The purpose of the current study was to examine clinicians' ability to apply the proposed ICD-11 diagnostic guidelines for schizoaffective disorder to case vignettes. To accomplish this goal, we compared the current ICD-10 guidelines to the proposed ICD-11 guidelines to determine if the new guidelines could be applied with at least similar accuracy. Many of the proposed changes in the guidelines aimed to increase clinical utility by removing potentially confusing or overlapping diagnoses and providing more information for each diagnosis (Gaebel et al., 2012). Based on the results from this first phase of the field trials,

**Table 2**
Comparison 1 diagnoses: schizophrenia vs. schizoaffective disorder.

| | ICD-11 | | ICD-10 | |
|---|---|---|---|---|
| | Vignette 1 | Vignette 2 | Vignette 1 | Vignette 2 |
| Schizophrenia | **101** | 37 | **116** | 64 |
| | **(72.14%)** | (26.43%) | **(71.60%)** | (39.51%) |
| Schizoaffective disorder | 12 | **62** | 8 | **46** |
| | (8.57%) | **(44.29%)** | (4.94%) | **(28.40%)** |
| Depressive episode with psychotic symptoms | 10 | 32 | 12 | 35 |
| | (7.14%) | (22.86%) | (7.41%) | (21.60%) |
| Another diagnosis | 17 | 9 | 26 | 17 |
| | (12.14%) | (6.43%) | (16.05%) | (10.49%) |

*Note:* bold denotes the correct option; $n = 140$ for ICD-11 and 162 for ICD-10.

**Table 3**
Comparison 2 diagnoses: schizoaffective disorder vs. depressive episode with psychotic symptoms.

| | ICD-11 | | ICD-10 | |
|---|---|---|---|---|
| | Vignette 2 | Vignette 3A | Vignette 2 | Vignette 3A |
| Schizophrenia | 47 (39.17%) | 0 (0%) | 73 (47.10%) | 0 (0%) |
| Schizoaffective disorder | **51** **(42.50%)** | 6 (5.00%) | **50** **(32.26%)** | 1 (0.65%) |
| Depressive episode with psychotic symptoms | 11 (9.17%) | **103** **(85.83%)** | 14 (9.03%) | **140** **(90.32%)** |
| Another diagnosis | 11 (9.17%) | 11 (9.17%) | 18 (11.61%) | 14 (9.03%) |

*Note:* bold denotes the correct option; $n = 120$ for ICD-11 and 155 for ICD-10.

clinicians applied the ICD-11 guidelines more accurately than the ICD-10 guidelines; however, the improvements were modest. ICD-11 guidelines include information related to differential diagnosis and additional features of the disorders. This additional information is not explicitly stated in ICD-10 and therefore could have contributed to ICD-11's improvements over ICD-10.

Although weak, diagnostic agreement for schizoaffective disorder in ICD-11 showed improvement over ICD-10. It is important to note that the kappas obtained in this study are not directly comparable to typical studies of diagnostic reliability, like the meta-analysis described in the introduction (Santelmann et al., 2015). Because of the differences in methodology (i.e., vignettes versus live patients), the reader should interpret the kappas from this study only in relative terms (i.e., improvement from ICD-10 to ICD-11 within this study). When examining the follow-up diagnostic questions, the most common discrepancy appeared to be related to the presence of mood symptoms in patients with clear symptoms of Schizophrenia. The vignettes were not likely the cause of the confusion because all of the vignettes were pretested and found to meet the correct diagnostic guidelines. Instead, the misdiagnoses and the corresponding responses to diagnostic questions could be the result of clinicians missing relevant information from the vignettes, falling prey to confirmatory biases, or actual confusion about the diagnostic guidelines.

Regarding the first explanation, it is possible that clinicians did not properly attend to the information in the vignette. However, lack of attention to details would not have resulted in the pattern of responses we observed, as they were selectively careless with one vignette and not others. More likely is the notion that the information about mood symptoms in Vignette 2 was overshadowed by the more salient psychotic symptoms. Diagnostic overshadowing has been shown to occur in a variety of areas (Alford and Locke, 1984; Reiss et al., 1982; Wood and Tracey, 2009), including with mood and psychotic symptoms overshadowing substance abuse problems (Goethe and Ahmadi, 1991; Skodol et al., 1984). However, we are not aware of any studies to date that explicitly examine psychotic symptoms overshadowing mood symptoms, as might have occurred in this study.

Second, the misdiagnoses could be due to clinicians' confirmatory biases. In other words, they may have started to formulate a diagnostic conclusion (e.g., Schizophrenia), which led them to discount or ignore other relevant symptoms (Garb, 1998). For example, in a sample of psychiatrists and medical students, 13% and 25% (respectively) engaged in confirmatory biases when searching for diagnostic information, leading to substantially increased chances of an incorrect diagnosis (Mendel et al., 2011).

Third, the confusion could be due to the guidelines themselves. Because the proposed revisions for the ICD are intended to increase the clinical utility of the diagnostic system, it will be important moving forward to identify ways that may aid clinicians' ability to accurately differentiate schizoaffective disorder from other psychotic and mood disorders. Although there is inherently going to be some level of overlap amongst specific psychotic and mood disorders due to the dimensional nature of the symptoms (Jonathan et al., 2013; Kendell and Jablensky, 2003; Malhi et al., 2008), it is still of the upmost importance to ensure that clinicians are able to meaningfully and reliably differentiate these diagnoses. ICD-11 has attempted to better differentiate schizoaffective disorder from Schizophrenia and mood disorders by providing more explicitly stated guidelines and differential diagnostic information. However, based on the present study, that additional information likely will not prove to be enough to substantially increase the diagnostic reliability of schizoaffective disorder. Lack of familiarity with the new guidelines may have contributed to the results, and clinicians may require additional education about the difference to help increase diagnostic reliability.

Last, the concept of schizoaffective disorder itself may be problematic. It exists as a practical solution to diagnosing individuals with overlapping psychotic and affective symptoms. Some solution is necessary, which is why the Working Group did not consider deleting schizoaffective disorder at this time. However, other solutions (like providing dimensional ratings) may capture the symptom picture more reliably than a categorical diagnosis. The inclusion in ICD-11 of separate dimensional ratings for depressive and manic mood symptoms in psychotic disorders would appear to be a step that direction (Keeley and

**Table 4**
Comparison 3 diagnoses: schizoaffective disorder vs. manic episode with psychotic symptoms.

| | ICD-11 | | ICD-10 | |
|---|---|---|---|---|
| | Vignette 2 | Vignette 3B | Vignette 2 | Vignette 3B |
| Schizophrenia | 47 (31.33%) | 2 (1.33%) | 57 (39.04%) | 2 (1.37%) |
| Schizoaffective disorder | **66** **(44.00%)** | 4 (2.67%) | **45** **(30.82%)** | 5 (3.42%) |
| Manic episode with psychotic symptoms | 0 (0%) | **122** **(81.33%)** | 1 (0.68%) | **128** **(87.67%)** |
| Depressive episode with psychotic symptoms | 29 (19.33%) | 0 (0%) | 29 (19.86%) | 0 (0%) |
| Another diagnosis | 8 (5.33%) | 22 (14.67%) | 14 (9.59%) | 11 (7.553%) |

*Note:* bold denotes the correct option; $n = 150$ for ICD-11 and 146 for ICD-10.

Gaebel, 2018). Nonetheless, the development of future classification systems may wish to revisit the conceptualization of schizoaffective disorder as a means of capturing combined mood and psychotic symptom presentations, possibly by implementing a fully dimensional scheme of psychotic and mood diagnoses if ratings prove to be clinically useful. Simply continuing to make iterative refinements in categorical guidelines for schizoaffective disorder is unlikely to produce substantial gains.

### 4.1. Limitations

By having such a large and diverse sample of participants, we would hope our findings would generalize to ICD-11 users in clinical practice. However, the majority of respondents were psychiatrists, which is not representative of mental health professionals generally (International Advisory Group, 2011). In higher-income countries, the proportion of psychiatrists to the general population is relatively high, but this distribution is not seen is lower-income countries which also utilize the ICD as their main diagnostic system (WHO, 2005).

Another limitation of this study that must be considered is the use of vignettes as opposed to live clinical cases. Vignettes were used to control for some of the inevitable variation in symptom presentations that would be present in a clinical population (Evans et al., 2015). However, in controlling for these variations, we limit our current findings to clear case presentations. Vignettes also limit the amount of information presented to participants, which would not likely be a problem found in a clinical setting. Specifically, the limited amount of information may alter a clinician's diagnosis or diagnostic considerations. Presenting only the prominent symptoms that would apply to a diagnosis may make diagnosing the vignette easier as compared to having an abundance of information that must be analyzed for accurate differential diagnoses in a clinical setting. A further limitation is that we only presented respondents with the diagnostic guidelines for psychotic disorders and select mood and stress disorders. Although we gave the clinicians in this study the option to write-in another diagnosis that was not listed, we did not include any of the diagnostic information for other potential diagnoses. Having these additional guidelines available may have increased the difficulty for clinicians to choose the accurate diagnosis because they would have had more options available. These limitations were addressed with additional clinic-based field trials in which clinicians interviewed patients to provide a diagnosis; kappas for Schizophrenia (0.87, 95%CI 0.84–0.89) and schizoaffective disorder (0.66, 95%CI 0.58–0.72) were higher for clinician interviews than in this study (Reed et al., 2018). This somewhat counterintuitive difference might be due to a clinicians' ability to ask clarifying questions in a live interview, which is not an option in a fixed vignette.

The final major limitation of this study is that we did not explicitly ask why clinicians were picking one diagnosis (correct or incorrect) over another diagnosis. For some of the comparisons being made, this differential diagnosis information could have been particularly helpful in determining which aspects of the diagnostic guidelines were aiding in a correct diagnosis or possibly negatively impacting respondents' ability to choose the correct diagnosis. Further analyses of the participants' responses could perhaps aid in generating ideas for why certain misdiagnoses were made while others were not.

#### Contributors

Jared Keeley, Wolfgang Gaebel, Juergen Zielasek, and Geoffrey Reed designed the study. Tahilia Rebello, Rebeca Robles, Chihiro Matsumoto, Cary Kogan, and Destiny Peterson oversaw data collection and implementation. Destiny Peterson and Jared Keeley conducted the statistical analysis. Destiny Peterson and Christopher Webb drafted the initial manuscript. All authors contributed to and have approved the final manuscript.

### References

Alford, J.D., Locke, B.J., 1984. Clinical responses to psychopathology of mentally retarded persons. Am. J. Ment. Defic. 89, 195–197.

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders. fifth edition. American Psychiatric Press, Washington, DC.

Evans, S.C., Roberts, M.C., Keeley, J.W., Blossom, J.B., Amaro, C.M., Garcia, A.M., Stough, C.O., Canter, K.S., Robles, R., Reed, G.M., 2015. Vignette methodologies for studying clinicians' decision-making: validity, utility, and application in ICD-11 field studies. Int. J. Clin. Health Psychol. 15, 160–170.

First, M.B., 2010. Clinical utility in the revision of the Diagnostic and Statistical Manual of Mental Disorders (DSM). Prof. Psychol. Res. Pract. 41, 465–473.

Gaebel, W., 2012. Status of psychotic disorders in ICD-11. Schizophr. Bull. 38, 895–898.

Gaebel, W., Zielasek, J., Cleveland, H.R., 2012. Classifying psychosis – challenges and opportunities. Int. Rev. Psychiatry 24 (6), 538–548. https://doi.org/10.3109/09540261.2012.737313.

Garb, H.N., 1998. Studying the Clinician: Judgment Research and Psychological Assessment. American Psychological Association, Washington, D.C.

Goethe, J.W., Ahmadi, K.S., 1991. Comparison of diagnostic interview schedule to psychiatrist diagnoses of alcohol use disorder in psychiatric inpatients. Am. J. Drug Alcohol Abuse 17, 61–69.

International Advisory Group for the Revisions of ICD-10 Mental and Behavioural Disorders, 2011. A conceptual Framework for the revision of the ICD-10 classification of mental and behavioural disorders. World Psychiatry 10, 86–92.

Jonathan, L.J., Chee, K., Ng, B., 2013. Schizoaffective disorder – an issue of diagnosis. ASEAN J. Psychiatry 14 (1), 76–81.

Keeley, J.W., Gaebel, W., 2018. Symptom rating scales for schizophrenia and other primary psychotic disorders in ICD-11. Epidemiol. Psychiatr. Sci. 27, 219–224. https://doi.org/10.1017/S2045796017000270.

Keeley, J.W., Reed, G.M., Roberts, M.C., Evans, S.C., Medina-Mora, M.E., Robles, R., Rebello, T., Sharan, P., Gureje, O., First, M.B., Andrews, H.F., Ayuso-Mateos, J.L., Gaebel, W., Zielasek, J., Saxena, S., 2016. Developing a science of clinical utility in diagnostic classification systems: field study strategies for ICD-11 mental and behavioural disorders. Am. Psychol. 71, 3–16. https://doi.org/10.1037/a0039972.

Kendell, R., Jablensky, A., 2003. Distinguishing between validity and utility of psychiatric diagnoses. Am. J. Psychiatry 160 (1), 4–12.

Malaspina, D., Owen, M.J., Heckers, S., Tandon, R., Bustillo, J., Schultz, S., Barch, D.M., Gaebel, W., Gur, R.E., Tsuang, M., Van Os, J., Carpenter, W., 2013. Schizoaffective disorder in DSM-5. Schizophr. Res. 150 (1), 21–25. https://doi.org/10.1016/j.schres.2013.04.026.

Malhi, G.S., Green, M., Fagiolini, A., Peselow, E.D., Kumari, V., 2008. Schizoaffective disorder: diagnostic issues and future recommendations. Bipolar Disord. 10, 215–230.

Mendel, R., Traut-Mattausch, E., Jonas, E., Leucht, S., Kane, J.M., Maino, K., Kissling, W., Hamann, J., 2011. Confirmation bias: why psychiatrists stick to wrong preliminary diagnoses. Psychol. Med. 41, 2651–2659. https://doi.org/10.1017/S0033291711000808.

Reed, G.M., 2010. Toward ICD-11: improving the clinical utility of WHO's international classification of mental disorders. Prof. Psychol. Res. Pract. 41, 457–464.

Reed, G.M., Rebello, T.J., Pike, K.M., Medina-Mora, M.E., Gureje, O., Zhao, M., Dai, Y., Roberts, M.C., Maruta, T., Matsumoto, C., Krasnov, V.N., Kulygina, M., Lovell, A.M., Stona, A.C., Sharan, P., Robles, R., Gaebel, W., Zielasek, J., Khoury, B., de Jesus Mari, J., Ayuso-Mateos, J.L., Evans, S.C., Kogan, C.S., Saxena, S., 2015. WHO's global clinical practice network for mental health. Lancet Psychiatry 2, 379–380. https://doi.org/10.1016/S2215-0366(15)00183-2.

Reed, G.M., Sharan, P., Rebello, T.J., Keeley, J.W., Medina-Mora, M.E., Gureje, O., Ayuso-Mateos, J.L., Kanba, S., Khoury, B., Kogan, C.S., Krasnov, V.N., Maj, M., Mari, J., Stein, D.J., Zhao, M., Akiyama, T., Andrews, H.F., Asevedo, E., Cheour, M., Dominguez-Martinez, T., El-Khoury, J., Fiorillo, A., Grenier, J., Gupta, N., Kola, L., Kulygina, M., Leal-Leturia, I., Luciano, M., Lusa, B., Nicolas, J., Martinez-Lopez, I., Matsumoto, C., Onofa, U.L., Paterniti, S., Purnima, S., Robles, R., Sahu, M.K., Sibeko, G., Zhong, N., First, M.B., Gaebel, W., Lovel, A.M., Maruta, T., Roberts, M.C., Pike, K.M., 2018. The ICD-11 developmental field study of reliability of diagnoses of high-burden mental disorders: results among adult patients in mental health settings of 13 countries. World Psychiatry 17, 174–186. https://doi.org/10.1002/wps.20524.

Reiss, S., Levitan, G.W., Szysko, J., 1982. Emotional disturbance and mental retardation: diagnostic overshadowing. Am. J. Ment. Defic. 86, 567–574.

Santelmann, H., Franklin, J., Bußhoff, J., Baethge, C., 2015. Test-retest reliability of schizoaffective disorder compared with schizophrenia, bipolar disorder, and unipolar disorder—a systematic review and meta-analysis. Bipolar Disord. 17, 753–768. https://doi.org/10.1111/bdj.12340.

Skodol, A.E., Williams, J.B.W., Spitzer, R.L., Gibbon, M., Kass, F., 1984. Identifying common errors in the use of DSM-III through diagnostic supervision. Hosp. Community Psychiatry 35, 251–255.

Wood, D.S., Tracey, T.J.G., 2009. A brief intervention for diagnostic overshadowing. Train Educ. Prof. Psychol. 3, 218–225. https://doi.org/10.1037/a0016577.

World Health Organization, 1992. The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines. World Health Organization, Geneva.

World Health Organization, 2005. Mental Health Atlas. Author, Geneva.