

PARAPHILIAS

Can Clinicians Use Dimensional Information to Make a Categorical Diagnosis of Paraphilic Disorders? An ICD-11 Field Study



Jared W. Keeley, PhD,¹ Peer Briken, MD,² Spencer C. Evans, PhD,³ Michael B. First, MD,^{4,5} Verena Klein, PhD,² Richard B. Krueger, MD,^{4,5} Chihiro Matsumoto, PhD,⁶ Ana Fresán, PhD,⁷ Tahlia J. Rebello, PhD,^{4,5} Rebeca Robles, PhD,⁷ Pratap Sharan, MD,⁸ and Geoffrey M. Reed, PhD^{4,9}

ABSTRACT

Background: The diagnosis of paraphilic disorder is a complicated clinical judgment based on the integration of information from multiple dimensions to arrive at a categorical (present/absent) conclusion. The recent update of the guidelines for paraphilic disorders in ICD-11 presents an opportunity to investigate how mental health professionals use the diagnostic guidelines to arrive at a diagnosis which thereby can optimize the guidelines for clinical use.

Aim: This study examined clinicians' ability to use the ICD-11 diagnostic guidelines for paraphilic disorders which contain multiple dimensions that must be simultaneously assessed to arrive at a diagnosis.

Methods: The study investigated the ability of 1,263 international clinicians to identify the dimensions of paraphilic disorder in the context of written case vignettes that varied on a single dimension only.

Outcomes: Participants provided diagnoses for the case vignettes along with dimensional ratings of the degree of presence of five dimensions of paraphilic disorder (arousal, consent, action, distress, and risk).

Results: Across a series of analyses, clinicians demonstrated a clear ability to recognize and appropriately integrate the dimensions of paraphilic disorders; however, there was some evidence that clinicians may over-diagnose non-pathological cases.

Clinical Translation: Clinicians would likely benefit from targeted training on the ICD-11 definition of paraphilic disorder and should be cautious of over-diagnosing.

Strengths and Limitations: This study represents a large international sample of health professionals and is the first to examine clinicians' ability to apply the ICD-11 diagnostic guidelines for paraphilic disorders. Important limitations include not generalizing to all clinicians and acknowledging that results may be different in direct clinical interactions vs written case vignettes.

Conclusion: These results indicate that clinicians appear capable of interpreting and implementing the diagnostic guidelines for paraphilic disorders in ICD-11. **Keeley JW, Briken P, Evans SC, et al. Can Clinicians Use Dimensional Information to Make a Categorical Diagnosis of Paraphilic Disorders? An ICD-11 Field Study. J Sex Med 2021;18:1592–1606.**

Copyright © 2021, International Society of Sexual Medicine. Published by Elsevier Inc. All rights reserved.

Key Words: Paraphilic Disorder; ICD-11; Diagnosis; Sensitivity & Specificity; Dimensions; Categories

Received March 12, 2021. Accepted June 23, 2021.

¹Virginia Commonwealth University, Department of Psychology, Richmond, VA, USA;

²Institute for Sex Research, Sexual Medicine and Forensic Psychiatry University Medical Center Hamburg-Eppendorf, Hamburg, Germany;

³University of Miami, Department of Psychology, Coral Gables, FL, USA;

⁴Columbia University Vagelos College of Physicians and Surgeons, Department of Psychiatry, New York, NY, USA;

⁵New York State Psychiatric Institute, New York, NY, USA;

⁶Japanese Society of Psychiatry and Neurology, Tokyo, Japan;

⁷National Institute of Psychiatry Ramon de la Fuente Muniz, Mexico City, Mexico;

⁸All India Institute of Medical Sciences, Department of Psychiatry, New Delhi, India;

⁹World Health Organization, Department of Mental Health and Substance Use, Geneva, Switzerland

Copyright © 2021, International Society of Sexual Medicine. Published by Elsevier Inc. All rights reserved.

<https://doi.org/10.1016/j.jsxm.2021.06.016>

The *International Classification of Diseases and Related Health Problems* is the official framework for the collection and reporting of health information by the 194 member countries of the World Health Organization (WHO). Its eleventh revision (ICD-11) was approved by the World Health Assembly in May 2019. Over the next several years, countries will implement changes in their health information systems, policies, and in some cases legislation as necessary for adoption and the implementation of the ICD-11. The ICD contains the classification of mental disorders that is most widely used around the world.¹ Currently, the version of the ICD most used by clinicians² is the Clinical Descriptions and Diagnostic Guidelines (CDDG)³ developed by the WHO Department of Mental Health and Substance Abuse to provide sufficiently detailed information for implementation in clinical settings. A similar version will be published for the ICD-11. The ICD-11 will directly affect users of the DSM, as the official US system for reporting and reimbursement is based on the ICD, and the ICD-11 will define the constituent categories and codes available for future use.⁴

As a part of the development of the ICD-11 CDDG,^{5,6} the WHO Department of Mental Health and Substance Abuse undertook a series of case-controlled internet-based field trials to evaluate the draft ICD-11 diagnostic guidelines.⁷ One major goal in developing the ICD-11 CDDG was to improve the clinical utility of the guidelines in order to contribute to the accurate identification and proper management of mental disorders, thereby decreasing the significant treatment gap and the global disease burden these conditions impose.⁴ An important component of the diagnostic guidelines is to provide clear, scientifically supported, and clinically useful information to clinicians about the boundary or threshold between normal variation and psychopathology.

The ICD-11 diagnostic guidelines for paraphilic disorders involve a number of significant changes relative to the ICD-10 CDDG, specifying a number of requirements that must be met in order for an atypical pattern of sexual interest to be diagnosable as a mental disorder.⁸ One goal of the changes was to ensure that foci of sexual attraction were not labeled as disordered simply because of their deviation from social or cultural norms. A paraphilic arousal pattern is a persistent and intense pattern of atypical (ie, statistically rare) sexual arousal manifested by sexual thoughts, fantasies, urges, and/or behaviors. Paraphilic arousal patterns are not in and of themselves considered to be mental disorders, as was the case in ICD-10.

In contrast, ICD-11 paraphilic disorders are limited primarily to patterns of sexual arousal involving non-consenting individuals whose age or status renders them unwilling or unable to consent (eg, children, an unsuspecting individual being viewed through a window). These include five specifically named paraphilic disorders in which the arousal pattern focuses on non-consenting individuals: Exhibitionistic Disorder, Voyeuristic Disorder, Pedophilic Disorder, Coercive Sexual Sadism Disorder and Frotteuristic Disorder. ICD-11 also provides a residual

category, Other Paraphilic Disorder Involving Non-Consenting Individuals, to be used for patterns of sexual arousal that meet the proposed general definition but are not sufficiently common to be individually named (eg, patterns of sexual arousal involving corpses or animals). A paraphilic disorder diagnosis involving non-consenting individuals requires that the paraphilic arousal pattern be sustained, focused, and intense as manifested by persistent sexual thoughts, fantasies, urges, or behaviors, and also that the person has either acted on the paraphilic urges, fantasies, or thoughts, or is markedly distressed by them.

Paraphilic arousal patterns that involve only solitary behavior or consenting individuals (eg, cross-dressing, fetishism, sexual masochism) can be diagnosed in ICD-11 as a Paraphilic Disorder Involving Solitary Behavior or Consenting Individuals only if the person is either (i) markedly distressed by the nature of the paraphilic arousal pattern and the distress is not simply a consequence of rejection or feared rejection of the arousal pattern by others; or (ii) the nature of the paraphilic behavior involves significant risk of injury or death either to the individual (eg, asphyxophilia or achieving sexual arousal by restriction of breathing) or to the partner (eg, consensual sadism that results in injuries requiring medical attention).

Based on these ICD-11 diagnostic requirements, clinicians making a diagnosis of a paraphilic disorder involving non-consenting individuals need to consider four dimensions when evaluating whether or not the diagnostic requirements are met. First, clinicians need to consider the extent to which the behavior reflects a sustained, focused, and intense pattern of atypical sexual arousal manifested by sexual thoughts, fantasies, urges, and/or behaviors (as opposed to behavior that is impulsive or opportunistic). At the extreme end of this dimension, some individuals are intensely focused on their paraphilic arousal pattern to the exclusion of other stimuli, whereas for others it may be a part of a larger sexual repertoire. Although it does convey information about the nature of the arousal pattern, preference for paraphilic sexual stimuli is not a requirement for the diagnosis. Another indicator of the intensity of the paraphilic arousal pattern may be the degree it is consistently experienced over time, in contrast to one that is only experienced intermittently. Second, clinicians must assess the extent to which the arousal pattern focuses on non-consenting individuals. For example, the fact that the other individual does not consent to various sexual acts plays a role in the arousal pattern of an individual with Coercive Sexual Sadism Disorder. This diagnosis is not intended to cover situations in which both individuals fully consent to participate in the behavior (eg, through role play). To qualify for a diagnosis of Coercive Sexual Sadism Disorder, the person must be specifically aroused by the suffering of a non-consenting individual, and the evidence a clinician may use to come to this determination can vary in severity and kind.⁸ Third, the diagnosing clinician must consider the degree to which the person has acted on the arousal pattern, in contrast to having sexual fantasies about acting or engaging in preparatory behaviors, which may not be evidence of a paraphilic

disorder. Fourth, the clinician must consider the extent to which the person is personally distressed by the arousal pattern, which also must be differentiated from distress-related fear of criminal consequences or social rejection. Fifth, in the case of paraphilic arousal patterns that involve solitary or consensual behavior, the clinician must consider the extent to which behavior related to the arousal pattern involves a significant risk of injury or death to the individual or to others.

This conceptualization of paraphilic disorders represents a major change from ICD-10, which primarily provides a description of the relevant paraphilic behavior, sometimes with the specification that the behavior needs to be recurrent or persistent and, in the case of pedophilia and sadomasochism, that it be preferred over other forms of sexual activity. In contrast, the ICD-11 requires the clinician to make a series of determinations that are then integrated into a complex whole. No one part is sufficient to make a diagnosis. The clinician must simultaneously gauge several relevant dimensions in order to come to a conclusion about whether the pattern of sexual interests and behaviors presented by the individual meets the diagnostic requirements for a paraphilic disorder.

CONSEQUENCES OF PARAPHILIC DISORDERS

The revised definition of paraphilic disorders in ICD-11 is partially prompted by the risk of a misdiagnosis. One major risk associated with the diagnosis of paraphilic disorders is over pathologizing sexual behaviors that do not fulfill the requirements of a mental disorder (ie, false positives). Paraphilic disorders that focus on non-consenting individuals may be associated with criminal sex offenses, although this is not necessarily the case. Even if they have never committed criminal acts, people with paraphilic disorders may be deeply stigmatized. For example, people with pedophilic disorder may be stigmatized to the point that other people wish they were dead.⁹ In the adjudication and penalization of sexual offenses, the potential legal consequences of a diagnosis can be dramatic, in part due to the widespread assumption that paraphilic arousal patterns cannot change.^{10,11} A paraphilic disorder diagnosis in this context can result in lifetime civil commitment or extensive and sometimes invasive mandatory treatment, as well as limitations on where one may live or work.¹² Therefore, the risks associated with a false-positive diagnosis may be higher than for many other disorders.

Therefore, the developers of the ICD-11 sought to distinguish between paraphilic disorders that are relevant to public health (eg, resulting in damage to the mental or physical health of others) and indicate the need for health services and those that are simply descriptions of private behaviors without appreciable public health impact and for which treatment is neither indicated nor sought.¹³ For this reason, the ICD-10 categories of fetishism, fetishistic transvestism (sic), and sadomasochism are no longer named paraphilic disorders in the ICD-11, although these arousal patterns may provide a partial basis for a diagnosis of

Other Paraphilic Disorder Involving Solitary Behavior or Consenting Individuals if the other diagnostic requirements are met.

Dimensional Information in Categorical Decisions

As described above, the diagnosis of paraphilic disorders represents a complex synthesis of information based on the clinicians' assessment of each component dimension of the diagnostic requirements for paraphilic disorders. Clinicians' decisions about whether or not to assign a diagnosis therefore represent a special case of using dimensional information in several domains to inform a single categorical decision. An important part of evaluating the diagnostic guidelines for paraphilic disorders is understanding whether practicing mental health professionals are able to use dimensional information as intended in making a diagnosis.

Using dimensional information to come to a categorical decision is not a new phenomenon; clinicians have been doing so as long as they have been making diagnoses.¹⁴ It is a recognition that there is value—that is, clinical utility—to carving a dimension at some point to create groups for a specific purpose. Several assessments of psychopathological symptoms and personality provide dimensional scores on a variety of domains that professionals interpret to inform diagnosis. A classic example is the determination of whether particular scores on tests of intellectual functioning and adaptive behavior, which are dimensional, are or are not indicative of a diagnosis of a disorder of intellectual development.¹⁵ The cut-point is arbitrary in the sense that there is no inherent demarcation in the distribution of scores that would justify a separate category. However, there is utility in that scores below that threshold predict other kinds of functional problems that can benefit from clinical attention.¹⁶

There has been growing recognition within the field of psychopathology of the value of dimensional constructs in psychiatric diagnosis.^{17,18} However, there has been concern that dimensional models may be too complicated.^{19,20} Given the complexity of five component dimensions within the diagnostic requirements for paraphilic disorders in ICD-11, it is important to evaluate whether clinicians are able to perform this task.

Despite the evidence of the benefit of using dimensional information in diagnosis, to date there are no studies of how clinicians *integrate* different dimensions in categorical diagnosis, much less in the context of paraphilic disorder. We believe that more thoroughly investigating how mental health professionals use dimensional information is crucial to improving diagnostic accuracy, especially in a case like paraphilic disorder where the decision is complicated and may have profound implications.

Current Study

The purpose of this study was to examine clinicians' application of the ICD-11 diagnostic guidelines for paraphilic disorders to clinical material in the form of case vignettes that were systematically manipulated with respect to the presence and intensity of the dimensions described above. The study was designed to

examine how clinicians determine the threshold of disorder from non-disorder in the context of atypical patterns of sexual arousal, and whether the dimensions identified above are weighed appropriately in assigning ICD-11 paraphilic disorder diagnoses.

METHOD

Participants

Participants were 1,263 members of the Global Clinical Practice Network (GCPN). The GCPN is an international, multilingual network of mental health professionals established for the purpose of implementing ICD-11 field studies. Members were recruited through national and regional professional associations; international and national conferences in psychology, psychiatry, and related disciplines; professional listservs; and word-of-mouth. For more information on the history and development of the GCPN, see Reed et al.²¹ Participants were invited to complete the study if they indicated advanced or fluent proficiency in one of the languages of the study (English, Japanese, or Spanish) and if at the time of the study they spent some portion of their professional time seeing patients in a clinical setting or supervising

clinical services provided by others. At the time of the study (Oct. 2014 – Nov. 2015), 7,238 GCPN members met these criteria. Of those, 1,684 (23.3%) opened the study link, with 1,263 (75.0%; 17.4% of invited) completing the study. Only participants with complete data were used in further analyses. Male participants (27.1% response) were slightly more likely to participate in the study than female participants (21.7%; $\chi^2(1) = 29.66, P < .001$). Further, response rates of participants from the African (32.1%) and Eastern Mediterranean (28.1%) regions were somewhat elevated as compared to average participation across regions (24.6%; $\chi^2(8) = 21.67, P < .01$). There were no other differences in demographic characteristics between those who did and did not participate in the study. Characteristics of the sample are presented in Table 1.

Materials

All materials were developed in English and then forward and back translated into Japanese and Spanish according to best practice recommendations.^{7,22} The developers of the study created five sets of vignettes. Each set of vignettes related to one of the five previously discussed dimensions of paraphilic disorders, with

Table 1. Participant demographics

	English <i>f</i> (%)	Spanish <i>f</i> (%)	Japanese <i>f</i> (%)	Total <i>f</i> (%)
Region				
Africa	30 (3.5)	1 (0.4)	0	31 (2.5)
North America	181 (21.2)	0	0	181 (14.3)
South America	34 (4.0)	161 (69.4)	0	195 (15.4)
Eastern Mediterranean	49 (5.7)	0	0	49 (3.9)
Europe	360 (42.2)	70 (30.2)	0	430 (34.0)
Southeast Asia	120 (14.1)	0	0	120 (9.5)
Western Pacific-Asia	33 (3.9)	0	177 (99.4)	210 (16.6)
Western Pacific-Oceania	45 (5.3)	0	0	45 (3.6)
Other	1 (0.1)	0	1 (0.6)	2 (0.2)
Gender				
Male	514 (60.3)	125 (53.9)	144 (80.9)	783 (62.0)
Female	336 (39.4)	106 (45.7)	34 (19.1)	476 (37.7)
Profession				
Counseling	51 (6.0)	5 (2.2)	2 (1.1)	58 (4.6)
Medicine	430 (50.4)	78 (33.6)	147 (82.6)	655 (51.9)
Nursing	20 (2.3)	3 (1.3)	2 (1.1)	25 (2.0)
Psychology	245 (28.7)	125 (53.9)	23 (12.9)	393 (31.1)
Social work	47 (5.5)	1 (0.4)	0	48 (3.8)
Sex Therapy	6 (0.7)	0	0	6 (0.5)
Speech Therapy	2 (0.2)	0	0	2 (0.2)
Occupational Therapy	25 (2.9)	4 (1.7)	1 (0.6)	30 (2.4)
Certified Peer Support Worker	5 (0.6)	0	0	5 (0.4)
Other	22 (2.6)	16 (6.9)	3 (1.7)	41 (3.2)
Age <i>M</i> (<i>SD</i>)	47.70 (11.54)	47.28 (11.22)	47.09 (10.21)	47.54 (11.30)
Years of Experience <i>M</i> (<i>SD</i>)	15.43 (10.22)	17.98 (10.08)	14.57 (9.53)	15.78 (10.15)
Total <i>N</i>	853	232	178	1263

each containing four vignettes that varied across four levels of that particular dimension while keeping the level of all other dimensions constant as much as possible. The first set of vignettes varied according to the strength of the arousal pattern (*arousal dimension*, vignette series A) and depicted individuals with a frotteuristic interest. The second set of vignettes varied according to the extent to which the other person had consented to engage in the act (*consent dimension*, vignette series B) and involved a pattern of sexual sadism. The third set of vignettes varied according to the degree to which the person has acted on the arousal pattern (*action dimension*, vignette series C) and involved an exhibitionistic arousal pattern. The fourth set varied according to the amount of distress experienced by the person regarding a consensual paraphilic arousal pattern (*distress dimension*, vignette series D) and involved a fetishistic arousal pattern. The final set varied across the degree to which there was a significant risk of injury or death related to a consensual paraphilic behavior (*risk dimension*, vignette series E) and involved a masochistic arousal pattern.

For example, the four vignettes representing the *arousal* dimension altered the degree to which an individual had an intense and focused pattern of arousal regarding frotteuristic acts and fantasies. Vignette A4 represented an accidental contact and unintentional arousal of an individual that otherwise held normative sexual interests. Vignette A3 represented an opportunistic frotteur who did not have spontaneous frotteuristic fantasies. Vignette A2 depicted an individual who intentionally sought out opportunities to rub against individuals for sexual arousal but was not exclusive in his interest. Vignette A1 described an individual with a preferential and nearly exclusive, sustained, and intense pattern of arousal focused on frotteuristic acts. Beyond the manipulated content, the vignettes were as identical as possible. All vignettes depicted heterosexual males to avoid any interactive effects of gender or sexual orientation. The first vignette always met the full diagnostic requirements for a diagnosis, while the fourth never met the requirements. The middle vignettes' (ie, 2 & 3) diagnostic status was intentionally unclear to be better able to ascertain clinicians' decision-making process regarding the diagnostic threshold. The vignettes are available for review online at https://osf.io/r65f4/?view_only=68ae32d51e3c4fc98811386d09eac136.

All vignettes were pretested to ensure that they reflected the manipulation of the intended dimensions and did not vary on unintended dimensions. Twenty content experts completed a total of 63 ratings of the vignettes. Based upon these ratings, vignettes were either confirmed to vary only on the intended dimension or were altered to clarify the content.

Procedure

All study procedures were approved by the WHO Ethics Review Board as well as local institutional IRBs. Eligible GCPN members received an individualized email invitation to complete the study that contained a unique link to Qualtrics, the online

platform for the study. Participants were presented with informed consent information and asked to indicate their willingness to participate in the study. Participants next completed a set of questions confirming their eligibility to participate in the study and assessing their expertise regarding paraphilic disorders. They then reviewed the proposed ICD-11 diagnostic guidelines for all paraphilic disorders.

After the presentation of the diagnostic guidelines, participants were randomly assigned to view a total of five vignettes, one of the four vignettes from each dimension, in random order. That is, each participant saw a total of five vignettes, one from each of the five sets (A – E) in the study. The order in which the vignettes were presented, ie, which dimension they represented, was randomized. Participants saw one and only one vignette from each series. After reading the vignette, participants were asked if the individual would meet the diagnostic requirements for the general definition of paraphilic disorder, and if so, which specific paraphilic disorder. Participants then rated the severity of each of the five dimensions (*arousal*, *consent*, *action*, *distress*, and *risk*) using a visual analog scale ranging from 0 (very low) to 100 (very high). For vignettes where the participant rated any level of distress beyond 0, they also indicated on another visual analog scale the degree to which the distress was accounted for by rejection or feared rejection by others. Participants were able to review the vignettes and the diagnostic guidelines while they were completing all ratings. Participants then repeated the process for the four remaining vignettes. Participants rated all dimensions for all vignettes in order to determine the specificity of the manipulated dimension. After completion, all participants received an email certificate documenting their participation and thanking them. Participation in the study took approximately one hour (median = 60.42 min.)

RESULTS

Diagnostic Decisions

Participants' determinations of whether the individuals in the vignettes qualified for a diagnosis of a paraphilic disorder are listed in Table 2. Vignettes are ordered such that the first was the most severe indication of the dimension, the fourth the least. For most dimensions, the first two vignettes were overwhelmingly considered disordered. The third was more ambiguous, and the fourth was most often not disordered, although a substantial portion (around one quarter to one third) of participants continued to diagnose the least severe vignette as disordered. Vignettes from the *distress* dimension (series D) were the exception; even in their severe form, fewer participants considered the presentation of a fetish disordered, even though the first vignette would definitely qualify for a diagnosis under the ICD-11 definition (Other Paraphilic Disorder Involving Solitary Behavior or Consenting Individuals) because of the individual's level of distress about his arousal pattern.

Table 2. Participants' diagnostic decisions about vignettes across domains

	Disorder determination	
	Yes	No
Arousal Dimension: Frotteuristic arousal pattern		
Vignette A1. seeks opportunities, exclusive focus	305 (95.31%)	15 (4.69%)
Vignette A2. seeks opportunities but not exclusive focus	302 (95.27%)	15 (4.73%)
Vignette A3. opportunistic without spontaneous fantasies	230 (73.02%)	85 (26.98%)
Vignette A4. accidental/unintentional	116 (37.30%)	195 (62.70%)
Consent Dimension: Sadistic arousal pattern		
Vignette B1. imposes sadistic acts with no consent	288 (89.16%)	35 (10.84%)
Vignette B2. imposes sadistic acts with assumed consent	266 (84.18%)	50 (15.82%)
Vignette B3. partner consents to some acts but not others	179 (58.31%)	128 (41.69%)
Vignette B4. partner gives with full, explicit consent	83 (26.18%)	234 (73.82%)
Action Dimension: Exhibitionistic arousal pattern		
Vignette C1. regular planning and execution with no intent of stopping	304 (96.82%)	10 (3.18%)
Vignette C2. regular planning, only acted on it once opportunistically	267 (86.69%)	41 (13.31%)
Vignette C3. has not acted but says he might	233 (70.61%)	97 (29.39%)
Vignette C4. has not acted and is confident he will not	112 (36.01%)	199 (63.99%)
Distress Dimension: Fetishistic arousal pattern		
Vignette D1. very distressed and wants to change himself	225 (69.01%)	101 (30.99%)
Vignette D2. ambivalent; is distressed but also enjoys the fetish and is not willing to give it up	200 (61.54%)	125 (38.46%)
Vignette D3. distressed only about social rejection	70 (22.80%)	237 (77.20%)
Vignette D4. not distressed	45 (14.75%)	260 (85.25%)
Risk Dimension: Masochistic arousal pattern		
Vignette E1. lost consciousness from asphyxiation, possible brain damage, but expressed desire to do it again	249 (84.41%)	46 (15.59%)

(continued)

Table 2. Continued

	Disorder determination	
	Yes	No
Vignette E2. left bruises but conscious after asphyxiation, aroused by the possibility of losing control	246 (77.12%)	73 (22.88%)
Vignette E3. superficial marks after asphyxiation, heeds warning not to escalate	105 (34.77%)	197 (65.23%)
Vignette E4. uses safe word to avoid harm, no escalation	88 (25.36%)	259 (74.64%)

Influence of Domains on Diagnostic Decisions

The next series of analyses focused on whether clinicians used the dimensions as intended in determining if a diagnosis of a paraphilic disorder should be assigned to the individual described in the vignette.

Series A: Arousal dimension. The mean ratings for the arousal dimension across the four series A (frotteurism) vignettes followed the expected pattern (Table 3). First, participants made increasingly higher ratings of the extent to which the vignette reflected a sustained, focused, and intense pattern of atypical sexual arousal across the four vignettes, and each vignette was statistically different from the others. Second, we examined the category response curve for each of the dimension ratings of the arousal vignettes, shown in Figure 1. The curves show a general gradation of the intensity of arousal ratings across the vignettes, such that the peaks of the curves shift towards lower values of arousal for each successive vignette. However, there is poor differentiation of the curves for Vignettes A1, A2, and A3, such that they appear to exist upon a continuous gradation of intensity (ie, the peaks do not emerge separately from each other). Only Vignette A4 is clearly operating upon a different response metric, indicating that participants only considered Vignette A4 to be a different class of individual, which is consistent with the intended construction of the vignette as clearly non-disordered. Third, we conducted a logistic regression of the five dimensions predicting clinicians' decision about whether or not the individual meets the diagnostic requirements for a paraphilic disorder. Results are presented in Table 4. As might be expected, arousal was the most important variable for predicting disorder decision of the arousal vignettes and was entered into the model first ($\chi^2(1) = 766.48$, $P < .001$, Nagelkerke $R^2 = .677$).

Series B: Consent dimension. Results indicate that clinicians were able to use the consent dimension as intended based on their responses to the series B (sexual sadism) vignettes.

Clinicians demonstrated a clear pattern of difference across their ratings of the degree to which the individual in the vignette was aroused by violation of consent (Table 3). The category response curve for consent demonstrated that participants differentiated each of the four vignettes in terms of the extent to which sexual thoughts, fantasies, urges and/or behaviors toward individuals who had not consented represented an integral feature of the sexual arousal pattern (see Figure S1 in the supplemental material online). However, while there was relatively strong agreement that this pattern was not present in vignette B4, participants were more variable in their rating of other vignettes where consent was less clear, demonstrated by platykurtic (relatively flat) curves (ie, the height of the line is modest and not asymptotic to the top of the scale). Vignette B3 evidenced a bimodal distribution where some participants did not see consent as an integral feature of the arousal pattern and others did, with relatively few participants falling in-between. In the logistic regression, consent was the most important variable for determining disorder status for series B ($\chi^2(1) = 424.63$, $P < .001$, Nagelkerke $R^2 = .393$; Table 4).

Series C: Action dimension. Clinicians also used the action variable as intended, with well differentiated ratings across the four series C (exhibitionism) vignettes of the extent to which the individual in the vignette had acted on his paraphilic arousal pattern (Table 3). The category response curves (Figure S2 online) support the notion that action was meaningfully differentiated among the vignettes, with each vignette following a different pattern. Action proved to be the most important variable in predicting clinicians' diagnostic decisions as well ($\chi^2(1) = 437.09$, $P < .001$, Nagelkerke $R^2 = .423$; Table 4).

Series D: Distress dimension. Clinicians appropriately differentiated the distress dimension, as demonstrated by their ratings of the extent to which the individual described in the series D (fetishism) vignettes was distressed about his sexual arousal pattern (Table 3). Further, the category response curve for distress was the only one that evidenced any degree of discrimination for these

Table 3. Mean (and SD) domain ratings for vignettes

Series A: Arousal					
	Vignette A1 n = 320	Vignette A2 n = 317	Vignette A3 n = 315	Vignette A4 n = 311	Test
Arousal	90.20 ^a (11.41)	83.30 ^b (15.16)	66.25 ^c (25.28)	42.58 ^d (33.08)	$F(3,1259) = 272.19, P < .001$
Consent	91.09 ^a (13.93)	85.85 ^b (18.65)	75.17 ^c (26.46)	49.08 ^d (35.98)	$F(3,1259) = 174.06, P < .001$
Action	90.97 ^a (13.11)	87.75 ^a (13.07)	76.36 ^b (20.98)	46.91 ^c (33.89)	$F(3,1259) = 263.51, P < .001$
Distress	38.65 ^a (30.01)	38.99 ^a (28.90)	40.37 ^a (29.32)	43.74 ^a (30.34)	$F(3,1259) = 1.93, ns$
Risk	18.60 ^a (25.48)	18.09 ^a (24.25)	14.47 ^a (22.33)	9.37 ^b (17.37)	$F(3,1259) = 11.15, P < .001$
Series B: Consent					
	Vignette B1 n = 323	Vignette B2 n = 316	Vignette B3 n = 307	Vignette B4 n = 317	Test
Arousal	83.69 ^a (15.38)	82.47 ^a (17.64)	78.49 ^b (19.50)	76.18 ^b (23.03)	$F(3,1259) = 10.63, P < .001$
Consent	77.54 ^a (22.14)	67.16 ^b (28.47)	44.08 ^c (33.58)	19.34 ^d (26.68)	$F(3,1259) = 273.49, P < .001$
Action	79.44 ^a (20.44)	78.63 ^a (21.51)	75.77 ^a (22.51)	75.93 ^a (22.25)	$F(3,1259) = 2.35, ns$
Distress	35.64 ^a (30.34)	32.78 ^{ab} (29.09)	34.96 ^{ab} (30.79)	29.40 ^b (28.02)	$F(3,1259) = 2.86, P < .05$
Risk	62.69 ^a (27.08)	63.12 ^a (28.27)	44.37 ^b (31.85)	29.50 ^c (28.09)	$F(3,1259) = 99.87, P < .001$
Series C: Action					
	Vignette C1 n = 314	Vignette C2 n = 308	Vignette C3 n = 330	Vignette C4 n = 311	Test
Arousal	84.72 ^a (15.12)	70.42 ^b (23.30)	64.65 ^c (26.30)	66.56 ^{bc} (23.67)	$F(3,1259) = 51.39, P < .001$
Consent	87.91 ^a (16.86)	77.31 ^b (23.45)	71.65 ^c (26.51)	57.99 ^d (34.37)	$F(3,1259) = 71.45, P < .001$
Action	86.99 ^a (15.00)	69.14 ^b (23.94)	46.93 ^c (31.30)	26.13 ^d (29.78)	$F(3,1259) = 327.19, P < .001$
Distress	40.09 ^a (28.45)	54.21 ^b (26.14)	45.93 ^c (27.57)	40.97 ^{ac} (29.14)	$F(3,1259) = 16.74, P < .001$
Risk	19.49 ^a (24.89)	17.45 ^a (22.23)	12.69 ^b (20.25)	8.12 ^c (15.92)	$F(3,1259) = 18.19, P < .001$
Series D: Distress					
	Vignette D1 n = 326	Vignette D2 n = 325	Vignette D3 n = 307	Vignette D4 n = 305	Test
Arousal	82.59 ^a (17.72)	77.68 ^b (20.12)	69.28 ^c (26.11)	70.37 ^c (27.41)	$F(3,1259) = 23.57, P < .001$
Consent	27.28 ^a (33.06)	17.82 ^b (29.02)	17.51 ^b (28.32)	13.75 ^b (26.43)	$F(3,1259) = 12.30, P < .001$
Action	77.59 ^a (18.41)	82.03 ^b (15.78)	80.87 ^{ab} (19.69)	83.72 ^b (18.66)	$F(3,1259) = 6.47, P < .001$
Distress	83.60 ^a (19.34)	73.98 ^b (21.82)	27.14 ^c (25.85)	6.15 ^d (14.56)	$F(3,1259) = 1001.38, P < .001$
Risk	6.77 ^a (14.35)	3.91 ^b (9.76)	4.29 ^b (11.43)	3.39 ^b (10.63)	$F(3,1259) = 5.28, P < .01$
Series E: Risk					
	Vignette E1 n = 295	Vignette E2 n = 319	Vignette E3 n = 302	Vignette E4 n = 347	Test
Arousal	81.64 ^a (19.22)	79.81 ^a (18.78)	67.65 ^b (25.55)	66.94 ^b (24.50)	$F(3,1259) = 38.56, P < .001$
Consent	21.59 ^a (32.38)	21.93 ^a (31.53)	17.16 ^{ab} (28.91)	15.14 ^b (25.08)	$F(3,1259) = 4.18, P < .01$
Action	84.08 ^a (18.27)	81.67 ^a (17.27)	72.65 ^b (23.77)	71.92 ^b (23.19)	$F(3,1259) = 27.68, P < .001$
Distress	26.11 ^a (27.34)	27.64 ^a (27.86)	23.21 ^{ab} (24.06)	19.61 ^b (23.64)	$F(3,1259) = 6.24, P < .001$
Risk	89.27 ^a (13.22)	77.27 ^b (21.91)	42.98 ^c (31.36)	24.58 ^d (25.26)	$F(3,1259) = 501.98, P < .001$

Note: Same superscripts across a row denote equal means.

vignettes (Figure S3 online); the curves for other dimensions did not differentiate among the vignettes. In the logistic regression, distress was the best predictor of diagnostic assignment ($\chi^2(1) = 356.20, P < .001$, Nagelkerke $R^2 = .330$; Table 4).

Series E: Risk dimension. Finally, clinicians also differentiated appropriately the extent to which there was a risk of injury or death (to either party) due to behavior related to the arousal pattern of the person described in the series E (sexual masochism)

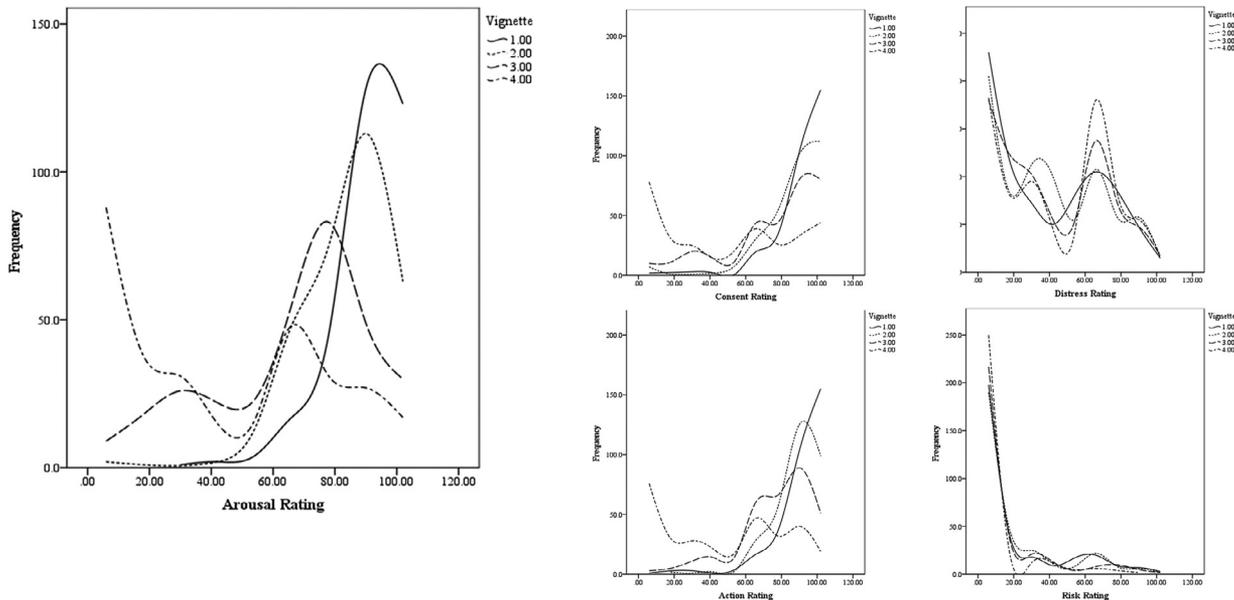


Figure 1. Category response curves for the arousal dimension.

vignettes. The risk dimension showed statistically significant differences across all means (Table 3). The category response curves for risk were also well differentiated across the vignettes (see Figure S4 online). Risk proved to be the best predictor of disorder status in the logistic regression ($\chi^2(1) = 544.44, P < .001$, Nagelkerke $R^2 = .468$; Table 4).

Confounding Influences on Diagnostic Decisions

While the manipulated dimensions each had the desired effect on clinicians' diagnostic decisions, the vignettes were designed to vary on one dimension only. Thus, if the dimensions of arousal, consent, action, distress, and risk were perceived by clinicians as being independent of one another, one would expect the ratings on dimensions not meant to vary to have equivalent ratings. However, if the rating of the other dimensions significantly impacted clinicians' ratings of the vignettes and diagnostic decisions besides the manipulated variable, it is evidence of an interaction among the domains that may be a function of clinicians' appraisal of the case (ie, their cognitive processes).

Series A: Arousal dimension. First, ratings of the arousal dimension (series A, frotteurism) were substantially correlated with ratings of consent ($r = 0.73$) and action ($r = 0.80$) while being minimally related to distress ($r = 0.04$) and risk ($r = 0.20$). The category response curves for each of the other dimensions indicate that consent and action followed the same basic response pattern as did arousal, further evidencing that participants conflated these variables in their understanding of the vignettes in series A (Figure 1). However, participants followed a different pattern for distress and risk, where participants did not differentiate the four vignettes (as intended). Further evidence for additional impact upon clinicians' decisions also comes from the logistic regression analysis. As can be seen in Table 4, all of the

remaining four variables provided significant prediction of clinicians' diagnostic decisions for the vignettes in series A, beyond the effect of the arousal dimension. In the second step, action significantly improved the prediction ($\chi^2(1) = 47.06, P < .001$, Nagelkerke $R^2 = .707$). Consent was added in the third step ($\chi^2(1) = 13.57, P < .001$, Nagelkerke $R^2 = .715$), distress¹ was added in the fourth step ($\chi^2(1) = 7.33, P < .01$, Nagelkerke $R^2 = .720$), and risk in the final step ($\chi^2(1) = 4.84, P < .05$, Nagelkerke $R^2 = .723$). When demographic factors were included as control variables, clinician gender was a significant predictor ($\chi^2(1) = 5.96, P < .05$, Nagelkerke $R^2 = .726$) with male participants (76.8%) being more likely to give a diagnosis than female participants (73.1%). When gender was included in the model, the effect of risk became marginally significant.

Series B: Consent dimension. The consent dimension (series B, sexual sadism) was modestly related to the degree to which the acts of the person were perceived to be harmful to both parties (ie, risk $r = 0.52$), with only minimal overlap with the other dimensions (arousal $r = 0.27$, action $r = 0.20$, and distress $r = 0.15$). However, risk provides a limited amount of information about clinicians' diagnostic decisions, as the category response curves are almost entirely redundant (Figure S1 online). There does appear to be a bimodal distribution, where a portion of participants rated relatively low risk (≈ 0), and then a different group of participants indicated a moderate level of risk (≈ 70). After splitting the distribution of risk at a natural cut point of 50, those who gave low ratings of risk were less likely to label the person disordered across the series B vignettes ($G^2(4) = 167.50, P < .0001$). The category curves for arousal, action, and distress

¹Controlling for distress due to fear of rejection had no impact on analyses for the action domain.

Table 4. Logistic regression of disorder status for vignettes

	-2 Log likelihood	B	SE	OR
Series A: Arousal				
Step 1	641.20			
Arousal		-.083*	.005	.920
Step 2	594.14			
Arousal		-.064*	.005	.938
Action		-.035*	.005	.966
Step 3	580.57			
Arousal		-.058*	.005	.944
Action		-.029*	.005	.971
Consent		-.017*	.004	.984
Step 4	573.23			
Arousal		-.057*	.005	.944
Action		-.031*	.006	.969
Consent		-.016*	.004	.985
Distress		-.011†	.004	.989
Step 5	568.39			
Arousal		-.057*	.005	.945
Action		-.031*	.006	.970
Consent		-.015*	.004	.985
Distress		-.010‡	.004	.990
Risk		-.013‡	.006	.987
Control	562.40			
Arousal		-.057*	.005	.945
Action		-.032*	.006	.969
Consent		-.016*	.005	.984
Distress		-.010‡	.004	.990
Risk		-.011§	.006	.989
Clinician Gender		.561‡	.230	1.75
Series B: Consent				
Step 1	1216.86			
Consent		-.039*	.002	.962
Step 2	1069.10			
Consent		-.030*	.002	.971
Risk		-.031*	.003	.969
Step 3	1010.53			
Consent		-.030*	.003	.971
Risk		-.029*	.003	.971
Arousal		-.033*	.005	.968
Step 4	1000.33			
Consent		-.030*	.003	.971
Risk		-.028*	.003	.972
Arousal		-.034*	.005	.967
Distress		-.009†	.003	.991
Step 5	995.63			
Consent		-.030*	.003	.971
Risk		-.028*	.003	.973
Arousal		-.030*	.005	.971
Distress		-.009†	.003	.991
Action		-.009‡	.004	.991
Control	989.36			
Consent		-.030*	.003	.970
Risk		-.028*	.003	.973

(continued)

Table 4. Continued

	-2 Log likelihood	B	SE	OR
Arousal		-.030*	.005	.970
Distress		-.009†	.003	.991
Action		-.009‡	.004	.991
Profession		.104‡	.044	1.110
Series C: Action				
Step 1	1048.00			
Action		-.044*	.003	.957
Step 2	941.27			
Action		-.038*	.003	.963
Arousal		-.036*	.004	.965
Step 3	909.92			
Action		-.035*	.003	.965
Arousal		-.030*	.004	.971
Consent		-.017*	.003	.983
Step 4	892.22			
Action		-.035*	.003	.966
Arousal		-.029*	.004	.971
Consent		-.018*	.003	.983
Distress		-.013*	.003	.987
Series D: Distress				
Step 1	1368.08			
Distress		-.032*	.002	.969
Step 2	1308.35			
Distress		-.032*	.002	.968
Consent		-.018*	.002	.982
Step 3	1274.17			
Distress		-.030*	.002	.971
Consent		-.017*	.002	.983
Arousal		-.020*	.004	.981
Step 4	1254.26			
Distress		-.029*	.002	.972
Consent		-.013*	.003	.987
Arousal		-.023*	.004	.978
Risk		-.033*	.008	.968
Control	1228.76			
Distress		-.029*	.002	.971
Consent		-.014*	.003	.986
Arousal		-.023*	.004	.977
Risk		-.033*	.008	.967
Clinician Gender		.559*	.146	1.748
Language		.305†	.100	1.356
Series E: Risk				
Step 1	1196.33			
Risk		-.045*	.002	.956
Step 2	1153.64			
Risk		-.044*	.002	.957
Distress		-.020*	.003	.980
Step 3	1112.46			
Risk		-.039*	.002	.961
Distress		-.021*	.003	.979
Arousal		-.024*	.004	.976

(continued)

Table 4. Continued

	-2 Log likelihood	B	SE	OR
Step 4	1106.78			
Risk		-.039*	.002	.962
Distress		-.019*	.003	.981
Arousal		-.023*	.004	.977
Consent		-.007 [‡]	.003	.993
Control	1086.05			
Risk		-.041*		.960
Distress		-.018*		.982
Arousal		-.023*		.977
Consent		-.006 [‡]		.994
Clinician Gender		.700*		2.014

* $P < .001$.[†] $P < .01$.[‡] $P < .05$.[§] $P < .10$.

B = unstandardized regression coefficient; OR = Odds Ratio; SE = Standard Error.

are largely redundant, indicating no differentiation of the vignettes on these dimensions that were not intended to vary. In the logistic regression, risk significantly improved the prediction beyond what was accounted for by consent ($\chi^2(1) = 147.76$, $P < .001$, Nagelkerke $R^2 = .501$). Arousal ($\chi^2(1) = 58.57$, $P < .001$, Nagelkerke $R^2 = .541$), distress² ($\chi^2(1) = 10.20$, $P < .01$, Nagelkerke $R^2 = .547$), and action ($\chi^2(1) = 4.70$, $P < .05$, Nagelkerke $R^2 = .550$) added significant portions of variability in the remaining steps, respectively. The clinicians' profession was the only demographic factor that predicted diagnosis ($\chi^2(1) = 5.44$, $P < .05$, Nagelkerke $R^2 = .554$); when it was in the model, there were no substantive changes to the effect of the other variables. Medical professionals (69.3%) were more likely to give a diagnosis than psychologists (58.3%) or other professionals (61.9%).

Series C: Action dimension. Other variables also appeared to influence clinicians' use of the action dimension (vignette series C, exhibitionism). Ratings of the action dimension evidenced moderate relationships with arousal ($r = 0.52$) and consent ($r = 0.46$), with small relationships with distress ($r = 0.14$) and risk ($r = 0.24$). Clinicians perceived an increasing importance of the violation of consent as part of the arousal pattern for vignettes where the person had been more likely to act upon his desires. When little action occurred (vignette C4; see Figure S2 online), clinicians were more willing to indicate that no violation of consent was part of the arousal pattern. In contrast, when the person had acted repeatedly upon his desires (vignette C1), clinicians perceived violation of consent to be more a part of his arousal pattern than for other vignettes (even though the relevant wording was the same). The relationship with ratings of arousal appears to be an effect of the least severe vignette (C4) being perceived by clinicians as having higher levels of arousal than the

other vignettes in the series. In the logistic regression, in addition to action, arousal ($\chi^2(1) = 106.73$, $P < .001$, Nagelkerke $R^2 = .506$), consent ($\chi^2(1) = 31.35$, $P < .001$, Nagelkerke $R^2 = .529$), and distress³ ($\chi^2(1) = 17.71$, $P < .001$, Nagelkerke $R^2 = .542$) each added significant predictive value to the equation. Risk was not a significant predictor in the model, nor was any demographic factor.

Series D: Distress dimension. Distress (series D, fetishism) appeared to have the least interference from any of the other dimensions. Correlations with each other dimension were small (arousal $r = 0.28$, consent $r = 0.13$, action $r = -0.01$, risk $r = 0.15$). Clinicians rated each vignette in series D in near-equivalent patterns for each non-manipulated variable (Figure S3 online). As such, it appears that clinicians were able to evaluate distress separately from the other variables in the definition of paraphilic disorder. However, somewhat counterintuitively, three of the remaining variables added significant variance to the logistic regression model: consent ($\chi^2(1) = 59.73$, $P < .001$, Nagelkerke $R^2 = .377$), arousal ($\chi^2(1) = 34.19$, $P < .001$, Nagelkerke $R^2 = .403$), and risk ($\chi^2(1) = 19.91$, $P < .001$, Nagelkerke $R^2 = .417$). Action did not add significantly to the prediction. Both clinician gender ($\chi^2(1) = 12.68$, $P < .001$) and language ($\chi^2(1) = 9.49$, $P < .01$, Nagelkerke $R^2 = .435$) were significant predictors when included as control variables. Male participants (46.4%) were more likely to give a diagnosis than female participants (37.0%). Participants that completed the study in Japanese (34.3%) were less likely to give a diagnosis than English-speaking (44.1%) or Spanish-speaking (44.4%) participants. The total amount of variability in the clinicians' decision accounted for by the series D model was somewhat weaker than for other series. Thus, it is possible that each dimension has a higher chance of representing unique variance, because there is more remaining variability to represent (vs overlapping with variability already accounted for by another dimension).

Series E: Risk dimension. Finally, the risk dimension (series E, sexual masochism) was moderately related to the arousal dimension ($r = 0.45$) with small relationships to the other dimensions (consent $r = 0.19$, action $r = 0.37$, distress $r = 0.20$). The category response curves (Figure S4 online) confirm that clinicians' ratings of the non-manipulated variables generally did not differ across vignettes. Clinicians perceived a slightly lower focus on the arousal pattern for the least severe vignette (E4). Three other variables were added to the logistic regression model: distress ($\chi^2(1) = 42.68$, $P < .001$, Nagelkerke $R^2 = .497$), arousal ($\chi^2(1) = 41.18$, $P < .001$, Nagelkerke $R^2 = .524$), and consent ($\chi^2(1) = 5.69$, $P < .05$, Nagelkerke $R^2 = .528$) in that order. Clinician gender was also a significant predictor ($\chi^2(1) = 19.81$, $P < .001$, Nagelkerke $R^2 = .540$), but it did not impact the effect of

²Controlling for distress due to fear of rejection led this variable to be dropped from the model.

³Controlling for distress due to fear of rejection led to minimal changes in the values reported here and the corresponding table. The interpretation and statistical significance remained unchanged.

the other variables when included as a control. Male participants (58.4%) were more likely to provide a diagnosis than female participants (47.9%). Because clinicians' ratings of distress could be due to the person fearing social rejection, we examined the degree to which clinicians' ratings of perceived social rejection accounted for their initial ratings of distress. Ratings of social rejection fully accounted for perceived distress impacting clinicians' diagnostic decision in the logistic regression; when it was included in the model, distress was no longer a significant predictor. Overall, clinicians were able to use the risk dimension independently of other dimensions in arriving at a diagnosis of paraphilic disorder.

DISCUSSION

Applying the ICD-11 guidelines to make a diagnosis of a paraphilic disorder is a complex decision in which several components must be considered simultaneously. The diagnostic decision encompasses dimensions of *arousal*, *consent*, *action*, *distress*, and *risk*. In this study, clinicians were able to clearly differentiate cases that varied along those dimensions. Despite concerns that integrating information from multiple dimensions might be too cognitively demanding a task, clinicians did well overall in integrating the different components of the diagnostic guidelines. Each dimension emerged as the best predictor of a categorical diagnosis for its respective series of vignettes. Additionally, the line at which most clinicians drew the distinction between disorder and non-disorder for each series of vignettes was close to what would be expected from the ICD-11 guidelines and what was confirmed by pretesting experts. There were no differences by profession other than for the *consent* dimension.

However, this study highlights the importance of empirically evaluating this decision-making process, as there was also evidence of overdiagnosis of paraphilic disorder to cases where it was not warranted. Between one-quarter and one-third of participating clinicians assigned a paraphilic disorder diagnosis to cases that should clearly have been considered non-disordered (the fourth vignette in all domains) according to the diagnostic guidelines. The high rate of false positive diagnoses may be related to previous diagnostic practice or differing views of non-normative sexual behavior that influence clinicians' diagnostic decisions in assigning paraphilic disorder diagnosis. It is also possible that clinicians did not read the diagnostic guidelines carefully enough and simply assigned the diagnosis based on the type of sexual behavior involved. Determining if these factors are in play would be an important focus of future research. Regardless, educational programs related to the implementation of ICD-11 should focus on clarifying the distinction between disorder and non-disorder in the context of paraphilic arousal patterns.

Some of the dimensions interacted in predicting a diagnosis. This finding illustrates the complexity of the diagnostic process. It is also important to note that the independence of the dimensions within this study was imposed by the study methodology

and not a reflection of clinical reality. These variables likely covary in clinical cases of paraphilic disorder, and patterns of covariance may differ across specific paraphilic disorders. Indeed, the ICD-11 guidelines imply some types of overlap, as engaging in the behavior (ie, *action*) is one of the important ways in which an arousal pattern can be manifested. The interactions we observed in this study may suggest that clinicians are also making use of these relationships in diagnostic decisions, and that this affects how clinicians perceive and conceptualize information from cases. On the other hand, the overlap may suggest that if one dimension is present, clinicians may infer the presence of other components of paraphilic disorders even when they are not explicitly present. That interactive effect may increase the risk of false-positive diagnoses.

The interactions that occurred among the dimensions may provide insight into clinicians' thinking about the components of the diagnostic guidelines for paraphilic disorders. First, when levels of arousal varied, clinicians perceived differences regarding how much the person had acted upon the interest and the degree to which consent was violated. This conflation is not entirely unexplainable, as the intentionality implied by consent would be partially measured by the individuals' willingness to act (or having acted) in violation of the consent of the other person. For example, the least severe vignette's engagement in the act was accidental. Because he did not intend to do it, participants may have reasonably viewed him as not having violated the consent of the person with whom he came into contact. When the person is perceived as willfully engaging in the behavior, the consent of the victim may be seen as more violated than in accidental behavior, and if the behavior is repetitive or the arousal pattern exclusive, as in vignette A1, then this may be perceived as constituting stronger consent violations. Because the guidelines themselves include some overlap between action and arousal, the vignettes may have not been entirely independent in this regard, despite expert pretesting confirming that they were. Additionally, it is possible that knowing someone has done something more often leads to the perception that he is more aroused by it. This pattern likely also explains the impact on ratings of arousal and consent when action was varied (ie, the reciprocal direction). Further, when consent was varied, clinicians perceived greater risk of harm. Again, this conflation is understandable, as a greater focus on the violation of consent implies higher potential of harm to the other party.

To our knowledge, this field trial represents the first attempt to investigate empirically the development of diagnostic guidelines for paraphilic disorders. The ICD-11 field trial process has been explicitly designed to ensure that feedback from the field trials occurs in time to make substantive changes as necessary to the diagnostic guidelines before their final release.⁷ The careful development of the diagnostic guidelines for paraphilic disorder is of great importance because of the sometimes dramatic consequences that false positive or negative diagnoses can have.^{9,23,24}

The fact that the majority of the clinicians followed the diagnostic guidelines correctly is encouraging for their eventual adoption in clinical settings. Nonetheless, there was a degree of overdiagnosis of non-pathological cases in the study, occasionally associated with demographic factors of the clinician like being male or a physician. These differences may imply that certain groups of clinicians are more prone to biased beliefs about sexual behavior, but that conclusion needs further investigation in future studies. If that is the case, targeting training of the diagnostic guidelines in connection with information and education about the large range of variation of healthy sexual fantasies and behaviors to groups more prone to bias should be helpful to improve diagnostic outcomes as well as reduce stigma among professionals in the future. The ICD-10 offered little guidance about differentiating variations in sexual behavior from paraphilic disorders; some of the overdiagnosis that occurred in this study may have been a result of clinicians operating on their understanding of the ICD-10 definitions. Nonetheless, practicing mental health professionals should exercise caution when diagnosing paraphilic disorders in the recognition that false-positive diagnoses may be common.

Limitations

The results of this study should be considered in light of its limitations. First, the sample of clinicians included in this study should not be viewed as representative of all clinicians worldwide. Despite the unprecedented diversity of the sample in terms of profession, language, and world region, participants who volunteered to be part of the GCPN are likely different from other professionals, given that they are interested enough in classification to volunteer their valuable time to participate in studies of the ICD-11 development. As such, their motivation to learn the new diagnostic guidelines for paraphilic disorders and their understanding of issues pertinent to the diagnosis of mental disorders may exceed those of their peers. Similarly, the response rate was low but consistent with other studies in this series²⁵⁻²⁸ and higher than other studies of mental health professionals.^{29,30} Nonetheless, the participants in this study represent a possible self-selection bias, and findings may be substantially different with a different group of participants. Second, this study employed written case vignettes, which are not perfect approximations of live clinical diagnosis that occurs in most clinical settings. Nonetheless, the vignettes that were formulated by experienced clinicians provide an ideal proxy for manipulating key features of a case, and include rich clinical detail that is sufficient to study clinical decision-making.²² Indeed, vignette studies have been shown to predict actual clinical behavior.³¹ However, it is possible that clinicians might not follow the same pattern of diagnosis if presented with live cases.

The vignettes only included male heterosexual cases of patients presenting with possible paraphilic disorders in order to avoid biasing the results due to clinicians' biases about sexual behavior in women or sexual minorities. Another recent case

vignette study found that gender has a pronounced effect upon the diagnosis of paraphilic disorders,³² and paraphilic symptoms in women might be more common than previously thought.³³ As such, a future study may evaluate the diagnostic guidelines for effects due to gender or sexual orientation.

Another limitation is that each set of vignettes was designed for one specific diagnosis (frotteurism, sexual sadism, exhibitionism, etc.). It is possible that the identification of a paraphilic disorder may also be influenced by which of the five dimensions (*arousal, action, risk, etc.*) was manipulated in each clinical picture. For example, *risk* was assessed in a case related to a masochistic arousal pattern, but clinicians' responses and cognitive process may be different if *risk* was assessed on pedophilic disorder. Future work may wish to examine if these dimensions are perceived similarly or differently across various types of paraphilic disorder. Additionally, despite our best attempts to vary the vignettes only on a single dimension, which was confirmed by expert pretest, some vignettes may have unintentionally varied on more than one dimension, reflecting the inherent artificiality of trying to separate the components of this definition. As such, any conclusions about the meaning of any covariance across dimensions should be interpreted cautiously. Last, this study was cognitively demanding, requiring complex judgments across five vignettes. Although the vignettes were presented in counterbalanced order, it is possible that participants became fatigued across the study, which took a median of one hour.

CONCLUSION

This study provides initial evidence that clinicians may be able to use the proposed ICD-11 diagnostic guidelines for paraphilic disorders to differentiate appropriately between cases and non-cases of paraphilic disorders in response to case vignettes. Further, clinicians were able to use multiple dimensions simultaneously in arriving at a diagnostic determination, even given their complex and interactive nature. This study provides an important first step towards validation of the proposed ICD-11 diagnostic guidelines for paraphilic disorders and suggests that future work evaluating clinicians' biases regarding paraphilic disorders could help improve rates of false-positive diagnoses.

ACKNOWLEDGMENTS

The authors thank the following individuals for their assistance in translating study materials into Spanish and Japanese and for testing the survey: Iván Arango, Tecelli Domínguez, Hirohiko Harima, Katsuki Harima, Nicolás Martínez, Yukako Nakagami, Go Nishiyama, Francisco Páez, Kyohei Yamamoto, and Takeshi Yamasaki. Lastly, we thank the clinician members of the Global Clinical Practice Network who contributed their valuable time and expertise as participants.

Corresponding Author: Jared W. Keeley, PhD, Psychology Department, 806 West Franklin Street, Box 842018, Richmond,

VA 23284-2018, USA. Tel: (804) 828-6120; Fax: (804) 828-2237; E-mail: jwkeeley@vcu.edu

Conflict of Interest: The authors report no conflicts of interest.

Funding: None.

STATEMENT OF AUTHORSHIP

Statement of authorship: JWK, PB, SCE, MBF, RK, CM, TJR, RR, and PS were members of or consultants to the WHO International Advisory Group for the Revision of ICD-10 Mental and Behavioral Disorders, the Field Studies Coordination Group for ICD-11 Mental and Behavioral Disorders, and/or the ICD-11 Working Group on Sexual Disorders and Sexual Health. GMR was a member of the WHO Secretariat. This article represents the opinions of its authors and except as specifically indicated does not represent the policies or positions of the World Health Organization.

REFERENCES

1. Reed GM, Correia JM, Esparza P, et al. The WPA-WHO global survey of psychiatrists' uses and attitudes towards mental disorders classification. *World Psychiatry* 2011;10:118–131.
2. First MB, Rebello TJ, Keeley JW, et al. Do mental health professionals use diagnostic classifications the way we think they do? A global survey. *World Psychiatry* 2018;17:187–195.
3. World Health Organization. The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines.. Geneva, Switzerland: World Health Organization; 1992. Available at: <http://www.who.int/classifications/icd/en/bluebook.pdf>. Accessed August 4, 2021.
4. Reed GM. Toward ICD-11: improving the clinical utility of WHO's international classification of mental disorders. *Prof Psychol Res Pr*. 2010;41:457–464. doi: [10.1037/a0021701](https://doi.org/10.1037/a0021701).
5. First MB, Reed GM, Saxena S, et al. The development of the ICD-11 clinical descriptions and diagnostic guidelines for mental and behavioural disorders. *World Psychiatry* 2015;14:82–90. doi: [10.1002/wps.20189](https://doi.org/10.1002/wps.20189).
6. International Advisory Group for the Revision of ICD-10 Mental and Behavioural Disorders. A conceptual framework for the revision of the ICD-10 classification of mental and behavioural disorders. *World Psychiatry* 2011;10:86–92. doi: [10.1002/j.2051-5545.2011.tb00022.x](https://doi.org/10.1002/j.2051-5545.2011.tb00022.x).
7. Keeley JW, Reed GM, Roberts M, et al. Developing a science of clinical utility in diagnostic classification systems: field study strategies for ICD-11 mental and behavioural disorders. *Am Psychol* 2016;71:3–16.
8. Krueger RB, Reed GM, First MB, et al. Proposals for paraphilic disorders in the international classification of diseases and related health problems, eleventh revision (ICD-11). *Arch Sex Behav* 2017;46:1529–1545. doi: [10.1007/s10508-017-0944-2](https://doi.org/10.1007/s10508-017-0944-2).
9. Jahnke S, Imhoff R, Hoyer J. Stigmatization of people with pedophilia: two comparative surveys. *Arch Sex Behav* 2015;44:21–34. doi: [10.1007/s10508-014-0312-4](https://doi.org/10.1007/s10508-014-0312-4).
10. Grundmann D, Krupp J, Scherner G, et al. Stability of self-reported arousal to sexual fantasies involving children in a clinical sample of pedophiles and hebephiles. *Arch Sex Behav* 2016;45:1153–1162.
11. Seto MC. Is pedophilia a sexual orientation? *Arch Sex Behav* 2012;41:231–236. doi: [10.1007/s10508-011-9882-6](https://doi.org/10.1007/s10508-011-9882-6).
12. Briken P, Krueger RB. From atypical sexual interests to paraphilic disorders: the planned ICD revisions related to paraphilic disorder. *J Sex Med* 2018;15:807–808.
13. Reed GM, Drescher J, Krueger RB, et al. Disorders related to sexuality and gender identity in the ICD-11: revising the ICD-10 classification based on current scientific evidence, best clinical practices, and human rights considerations. *World Psychiatry* 2016;15:205–221.
14. Blashfield RK, Keeley JW, Flanagan EH, et al. The cycle of classification: DSM-I through DSM-5. *Annu Rev Clin Psychol* 2014;10:25–51.
15. Salvador-Carulla L, Reed GM, Vaez-Azizi LM, et al. Intellectual developmental disorders: towards a new name, definition and framework for “mental retardation/intellectual disability” in ICD-11. *World Psychiatry* 2011;10:175–180.
16. Tassé MJ, Balboni G, Navas P, et al. Identifying behavioral indicators for intellectual functioning and adaptive behavior for ICD-11 disorders of intellectual development. *J Intellect Disabil Res* 2019;63:386–407.
17. Hyman S. The diagnosis of mental disorders: the problem of reification. *Annu Rev Clin Psychol* 2010;6:155–179.
18. Widiger TA, Samuel DB. Diagnostic categories or dimensions: a question for DSM-V. *J Abnorm Psychol* 2005;114:494–504.
19. Rottman BM, Ahn WK, Sanislow CA, et al. Can clinicians recognize DSM-IV personality disorders from five-factor model descriptions of patient cases? *Am J Psychiatry* 2009;66:427–433.
20. Rottman BM, Kim NS, Ahn WK, et al. Can personality disorder experts recognize DSM-IV personality disorders from five-factor model descriptions of patient cases? *J Clin Psychiatry* 2011;72:630–639.
21. Reed GM, Rebello TJ, Pike KM, et al. WHO's global clinical practice network for mental health. *Lancet Psychiatry* 2015;2:379–380. doi: [10.1016/S2215-0366\(15\)00183-2](https://doi.org/10.1016/S2215-0366(15)00183-2).
22. Evans SC, Roberts MC, Keeley JW, et al. Vignette methodologies for studying clinicians' decision-making: validity, utility, and application in ICD-11 field studies. *Int J Clin Health Psychol* 2015;15:160–170. doi: [10.1016/j.ijchp.2014.12.001](https://doi.org/10.1016/j.ijchp.2014.12.001).
23. Frances AJ, Widiger T. Psychiatric diagnosis: lessons from the DSM-IV past and cautions for the DSM-5 future. *Annu Rev Clin Psychol* 2012;8:109–130. doi: [10.1146/annurev-clinpsy-032511-143102](https://doi.org/10.1146/annurev-clinpsy-032511-143102).
24. Wakefield JC. DSM-5 proposed diagnostic criteria for sexual paraphilias: tensions between diagnostic validity and forensic utility. *Int J Law Psychiatry* 2011;34:195–209.

25. Keeley JW, Reed GM, Roberts MC, et al. Disorders specifically associated with stress: a case-controlled field study for ICD-11 mental and behavioural disorders. *Int J Clin Hlth Psy* 2016;16:109–127.
26. Heyman RE, Kogan CS, Foran HM, et al. A case-controlled field study evaluating ICD-11 proposals for relational problems and intimate partner violence. *Int J Clin Hlth Psyc* 2018;18:113–123. doi: [10.1016/j.ijchp.2018.03.001](https://doi.org/10.1016/j.ijchp.2018.03.001).
27. Keeley JW, Gaebel W, First MB, et al. Psychotic disorder symptom rating scales: are dichotomous or multi-point scales more clinical useful?—An ICD-11 field study. *Schiz Res* 2018;202:254–259. doi: [10.1016/j.schres.2018.07.006](https://doi.org/10.1016/j.schres.2018.07.006).
28. Claudino A, Pike KM, Hay P, et al. The classification of feeding and eating disorders in the ICD-11: results of a field study comparing proposed ICD-11 guidelines with existing ICD-10 guidelines. *BMC Med* 2019;17:93. doi: [10.1186/s12916-019-1327-4](https://doi.org/10.1186/s12916-019-1327-4).
29. Keeley JW, DeLao CS, Kirk CL. The commutative property in comorbid diagnosis: does $A+B = B+A$? *Clin Psychol Sci*. 2013;1:16–29.
30. Webb CA, Keeley JW. Evaluating clinicians' representations of schizoaffective disorder. *Compr Psychiatry* 2017;74:102–108. doi: [10.1016/j.comppsy.2017.01.009](https://doi.org/10.1016/j.comppsy.2017.01.009).
31. Langlely GR, Tritchler DL, Llewellyn-Thomas HA, et al. Use of written cases to study factors associated with regional variations in referral rates. *J Clin Epidemiol* 1991;44:391–402. doi: [10.1016/0895-4356\(91\)90077-M](https://doi.org/10.1016/0895-4356(91)90077-M).
32. Fuss J, Briken P, Klein V. Gender bias in clinicians' pathologization of atypical sexuality: a randomized controlled trial with mental health professionals. *Sci Rep* 2018;8:3715. doi: [10.1038/s41598-018-22108-z](https://doi.org/10.1038/s41598-018-22108-z).
33. Baur E, Forsman M, Santtila P, et al. Paraphilic sexual interests and sexually coercive behavior: a population-based twin study. *Arch Sex Behav* 2016;45:1163–1172. doi: [10.1007/s10508-015-0674-2](https://doi.org/10.1007/s10508-015-0674-2).

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.jsxm.2021.06.016](https://doi.org/10.1016/j.jsxm.2021.06.016).