



International Journal of Clinical and Health Psychology

www.elsevier.es/ijchp



ARTÍCULO TEÓRICO

Vignette methodologies for studying clinicians' decision-making: Validity, utility, and application in ICD-11 field studies[☆]



Spencer C. Evans^a, Michael C. Roberts^{a,*}, Jared W. Keeley^b, Jennifer B. Blossom^a,
Christina M. Amaro^a, Andrea M. Garcia^a, Cathleen Odar Stough^a,
Kimberly S. Canter^a, Rebeca Robles^c, Geoffrey M. Reed^d

^a University of Kansas, USA

^b Mississippi State University, USA

^c Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz, Ministry of Health, Mexico

^d World Health Organization, Switzerland

Received 5 September 2014; accepted 15 December 2014

Available online 29 January 2015

KEYWORDS

Vignette methodology;
Experimental design;
Clinical decision-making;
International Classification of Diseases (ICD-11);
Theoretical study

Abstract Vignette-based methodologies are frequently used to examine judgments and decision-making processes, including clinical judgments made by health professionals. Concerns are sometimes raised that vignettes do not accurately reflect “real world” phenomena, and that this affects the validity of results and conclusions of these studies. This article provides an overview of the defining features, design variations, strengths, and weaknesses of vignette studies as a way of examining how health professionals form clinical judgments (e.g., assigning diagnoses, selecting treatments). As a “hybrid” of traditional survey and experimental methods, vignette studies can offer aspects of both the high internal validity of experiments and the high external validity of survey research in order to disentangle multiple predictors of clinician behavior. When vignette studies are well designed to test specific questions about judgments and decision-making, they can be highly generalizable to “real life” behavior, while overcoming the ethical, practical, and scientific limitations associated with alternative methods (e.g., observation, self-report, standardized patients, archival analysis). We conclude with

[☆] Five of the authors (SCE, MCR, JWK, RR, and GMR) are affiliated with the ICD-11 Mental and Behavioural Disorders Field Studies Coordination Group and the World Health Organization (WHO) Department of Mental Health and Substance Abuse. However, the views expressed in this article are those of the authors and, unless specifically stated, do not reflect the official policies or views of WHO.

* Corresponding author. Clinical Child Psychology Program, Dole Human Development Center, 1000 Sunnyside Avenue, University of Kansas, Lawrence, KS, 66045, USA.

E-mail address: mroberts@ku.edu (M.C. Roberts).

methodological recommendations and a description of how vignette methodologies are being used to investigate clinicians' diagnostic decisions in case-controlled field studies for the ICD-11 classification of mental and behavioural disorders, and how these studies illustrate the preceding concepts and recommendations

© 2014 Asociación Española de Psicología Conductual. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

PALABRAS CLAVE

Metodología de viñetas; diseño experimental; toma de decisiones clínicas; Clasificación Internacional de Enfermedades (CIE-11); estudio teórico

Metodología basada en viñetas para el estudio de toma de decisiones clínicas: validez, utilidad y aplicación en los estudios de campo de la CIE-11

Resumen Las metodologías basadas en viñetas se utilizan frecuentemente para examinar los procesos de toma de decisiones, incluyendo los de profesionales sanitarios. No obstante, existen dudas sobre si las viñetas reflejan adecuadamente los fenómenos del "mundo real" permitiendo resultados y conclusiones válidas. Ofrecemos una visión de las características, variaciones de diseño, fortalezas y debilidades de estos estudios para examinar cómo los profesionales forman juicios clínicos (como el diagnóstico y tratamiento). Siendo "híbridos" de las encuestas tradicionales y los métodos experimentales, estos estudios pueden ofrecer la alta validez interna de los experimentos y la alta validez externa de las encuestas, para aislar múltiples factores predictivos del comportamiento de los clínicos. Un diseño adecuado para poner a prueba preguntas específicas acerca de los juicios y la toma de decisiones permite resultados altamente generalizables a la "vida real", sin las limitaciones éticas, prácticas y científicas de los métodos alternativos (observación, auto-informe, pacientes estandarizados, análisis de archivos clínicos). Concluimos con recomendaciones metodológicas que se ilustran tras una descripción del uso de las metodologías de viñetas para investigar las decisiones diagnósticas de los clínicos en los estudios de campo de casos y controles para la clasificación de los trastornos mentales y del comportamiento en la CIE-11

© 2014 Asociación Española de Psicología Conductual. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Vignettes have a long history of use to investigate a number of phenomena in the social, behavioral, and health sciences (e.g., [Alexander & Becker, 1978](#); [Bachmann et al., 2008](#); [Wallander, 2009](#)). By allowing the investigator to manipulate specific aspects of a written stimulus while controlling others, vignette-based experimental designs offer a glimpse into how individuals' thoughts, feelings, behaviors, and decisions are affected by factors that may not be easily discernible in real-life situations because of confounding sources of variability that cannot be controlled. These types of questions are of great interest for research on health care service provision because the judgments and actions made by health professionals—and specific aspects thereof, such as quality of care or conformance with practice standards—have enormous implications for individual patients, for health systems, and for governments and other payers. However, it is often unethical or impossible to assess clinical decision-making experimentally with real clinicians and patients in health care settings.

As a solution to this conundrum, vignette designs may be an ideal method for investigating how health clinicians make decisions that affect their patients. However, vignette-based studies have also attracted some criticisms in regard to potential limitations in construct and external validity. Indeed, a key consideration, inherent to all vignette

studies, is the extent to which a written stimulus, and participants' responses to it, can accurately represent certain aspects of what happens in the "real world." These are critical questions that warrant careful attention in the design and evaluation of studies.

The aim of this paper is to review and clarify the methodological issues related to vignette-based designs for studying clinicians' decision-making processes and behaviors. We focus on clinical decision-making as a complex cognitive and behavioral process that involves the use of knowledge and skills acquired from one's professional training and experience, as applied to particular scenarios in which the clinician is required to make a decision that affects a patient (e.g., assigning diagnoses, selecting treatments). First, we provide an overview of the defining features of vignette-based methodologies followed by a selection of research examples from a range of disciplines. Next, we explore several variations of vignette-based designs. We discuss the strengths and limitations of vignette-based approaches, with particular attention to validity concerns and whether vignette studies can accurately simulate and generalize to the "real world." Although we consider research from a number of disciplines, our review focuses particularly on the application of vignette methodologies in studies among health professionals. We provide several

methodological recommendations to ensure the quality and generalizability of vignette-based research in this area. We conclude with a description of how vignettes are being used to examine clinicians' diagnostic decision-making processes as part of case-controlled field studies related to the development of the Mental and Behavioural Disorders chapter of the 11th revision of the International Classification of Diseases and Related Health Conditions (ICD-11; [World Health Organization \[WHO\], 2014](#)), and discuss the design of those studies in relation to the preceding methodological concepts and recommendations.

Defining features of vignette studies

A vignette¹ is a brief, carefully written description of a person or situation designed to simulate key features of a real world scenario (e.g., [Alexander & Becker, 1978](#); [Atzmüller & Steiner, 2010](#); [Gould, 1996](#); [Schoenberg & Ravidal, 2000](#)). The content and characteristics of a vignette can be described as consisting of up to three different aspects, based on their function in the study design: (a) *experimental* aspects, which are systematically manipulated across vignettes to assess their effect on the dependent variables; (b) *controlled* aspects, which are kept consistent (i.e., identical or similar) across vignettes in order to eliminate extraneous variance; and (c) in some cases, *contextual* aspects, which demonstrate some variation across vignettes in order to provide verisimilitude (e.g., nonessential details that enhance the "personhood" of a vignette character), but are not thought to exert a causal influence on the dependent variables.

By parsing the variance of vignette content into these three parts, researchers can strive toward overall *vignette equivalence*—where the text within a vignette is similarly apportioned across all vignettes in the study. Moreover, through careful attention to all aspects of the vignette, researchers can ensure that the vignettes are equivalent in all major respects, except for the careful and deliberate *manipulation of experimental variables*. In other words, vignette equivalence provides a consistent, controlled backdrop against which researchers can systematically vary key independent variables to address questions of interest, while also allowing for examination of differences among participants (e.g., country of residence, language, professional background). For example, [Ross, Moffat, McConnachie, and Wilson \(1999\)](#) used vignettes to study general practitioners' identification and management of depressive symptoms in a patient presenting for a non-psychological concern. The *experimental* aspects of the vignettes include the gender of the patient (M, F) and the severity of the depressive symptoms (mild, moderate, severe). The *controlled* aspects include the patient's age (22 years old), the nature of the presenting concern (mild acne), and the style of the vignettes (brief, ambiguous, and sequential with increasing severity). In this study,

there was no *contextual* variation because the experimental manipulation was performed through modifications to a single vignette, rather than through generating different vignettes.

Another defining feature of vignette studies is that they employ participant-reported data collection strategies (i.e., self-reported responses to questions about the vignette), which are characteristic of traditional survey methodologies. These strategies can include quantitative and/or qualitative approaches that provide standardized questions to be answered by all participants. Just as vignette equivalence ensures the consistency of the experimental stimuli, *response consistency*—where items and response options are standardized for all participants—ensures that the data gathered are interpretable in a consistent manner, allowing meaningful conclusions to be drawn (e.g., [Gould, 1996](#); [Ludwick et al., 2004](#)).

Historical background

Over the past half-century, vignettes have been used to address myriad questions across a wide range of scientific fields and professional disciplines, including business, marketing, and economics (e.g., [Wason, Polonsky, & Hyman, 2002](#)); social and experimental psychology (e.g., [Jones & Aronson, 1973](#)); sociology ([Wallander, 2009](#)); developmental psychology (e.g., [Howie, Nash, Kurukulasuriya, & Bowman, 2012](#)); and education and school psychology ([Baudson & Preckel, 2013](#)). Early vignette studies often used written hypothetical scenarios for non-experimental purposes, for example, in field research in anthropology ([Herskovitz, 1950](#)), and clinician research and assessment in medicine ([Williamson, 1965](#)). Vignettes continue to be used for assessment in mental health care, including diagnostic assessment in mental health care ([Barnhill, 2014](#)) and professional ethics in psychology and medicine ([Macpherson & Veatch, 2010](#); [McCarron & Stewart, 2011](#)). Social and experimental psychologists were among the first to utilize vignettes for experimental research, primarily to investigate factors affecting people's perceptions of guilt and responsibility in hypothetical cases of rape, murder, and car accidents (e.g., [Jones & Aronson, 1973](#); [McGlynn, Megaw, & Bensen, 1976](#); [Walster, 1966](#)).

Vignettes have been particularly useful for research in the health and mental health professions. In nursing and medicine, studies have examined professionals' attitudes, perceptions, and beliefs (e.g., [Hughes & Huby, 2001](#)), quality of care in outpatient settings (e.g., [Peabody, Luck, Glassman, Dresslhaus, & Lee, 2000](#)), forensic psychiatric judgments concerning criminal insanity (e.g., [Grøndal, Grønnerød, & Sexton, 2009](#)), and general practitioners' decision making (e.g., [Wainwright, Gallagher, Tompsett, & Atkins 2010](#)). In social work, researchers have used factorial surveys to examine how social workers make decisions about diagnosis and treatment ([Taylor, 2006](#); [Wallander, 2012](#)), as well as to understand how individuals view and make decisions about issues such as child protection, race, and poverty (e.g., [Stokes & Schmidt, 2011](#)). Similarly, in health and health psychology, vignettes have been utilized to assess perceptions and attitudes, such as children's perceptions of peers with chronic illnesses or behavioral disorders (e.g.,

¹ In this article we focus our definition of "vignette" on narrative text presentations only; we do not discuss other media formats (e.g., photographs, illustrations, video, audio, games), which could also be considered types of vignettes.

Levy Cole, Roberts, & McNeal, 1996; Maieron, Roberts, & Prentice-Dunn, 1996; Roberts, Johnson, & Beidleman, 1984).

Research design variations

Most vignette studies are characterized by the same basic logic and variations as other experimental designs (Finger & Rand, 2005), in which different levels in the independent variable are operationalized as different conditions (i.e., vignettes) with which participants may be presented. Major variations of this methodology include *between-subjects* designs, where participants are randomly assigned to different vignette groups, such that all participants within a group are presented with the same vignette and participants in different groups are presented with different vignettes; *within-subjects* designs, where all participants are presented with a set of multiple vignettes; or *mixed designs*, where participants are randomly assigned to different groups, and within each group participants are presented with different sets of vignettes (Atzmüller & Steiner, 2010). When used in vignette studies, each of these design variations carries with it strengths and weaknesses similar to those associated with other between-subject, within-subject, or mixed design studies (Atzmüller & Steiner, 2010; Finger & Rand, 2005). These traditional experimental methods continue to be employed frequently and are well suited for research questions where the independent variables are categorical and relatively few.

A somewhat newer design is the factorial survey (Rossi & Nock, 1982; Rossi, Sampson, Bose, Jasso, & Passel, 1974). Whereas traditional experimental designs require separate vignettes for each condition, factorial surveys allow the experimental variables (continuous and categorical) to be experimentally varied directly in the vignette for each participant (Ludwick et al., 2004; Taylor, 2006; Wallander, 2009, 2012). Returning to our earlier example vignette study (Ross et al., 1999), suppose the researchers were also interested in how other demographic characteristics of the patient (e.g., age, ethnicity, income level, marital status, number of children) would affect the general practitioners' assessment and treatment of the depressive symptoms. In a traditional experimental design, this question would entail a factorial structure where continuous variables (age, income level) would be converted into categorical factors, and the number of participant groups would increase exponentially with each additional factor, thus requiring an enormous sample size. In the factorial survey design, however, the different values of these variables (e.g., age [20, 25, 30 . . . 75], marital status [single, married]) can be randomly varied directly in the text of the vignette for each participant. Each vignette variation is uniquely and automatically generated from a very large number of possible variations—in this example, 6,912 possible vignettes—and might be seen by only one participant, or not seen at all.

The design and analyses in a factorial survey, then, are based on the patterns of variation associated with characteristics of the vignettes sampled (based on survey methods and regression/correlation) rather than comparison of differences associated with vignette conditions (based on experimental methods and ANOVA) (Ludwick et al., 2004; Taylor, 2006; Wallander, 2009, 2012). Thus, the factorial

design can be more efficient for research questions involving continuous variables and a large number of factors (e.g., Taylor, 2006). However, because this method can quickly increase the complexity of a design, it may not be appropriate for research questions that could be more parsimoniously addressed with a traditional experimental design (Atzmüller & Steiner, 2010; Sniderman & Grob, 1996).

The vignette world and the real world

Throughout the literature, numerous proponents (e.g., Alexander & Becker, 1978; Ludwick et al., 2004; Wallander, 2009, 2012) and critics (e.g., Cullen, 2010; Faia, 1980; Stolte, 1994) of vignette research have echoed concerns about the artificiality of vignettes. That is, textual descriptions and hypothetical behavior might not be sufficiently representative of real-world phenomena, and this raises concerns about the validity of research findings and conclusions based on them (see also Gould, 1996; Hughes, 1998; Hughes & Huby, 2001; Kim, 2012; Spalding & Phillips, 2007).

To accurately understand and address these concerns, it is necessary to first consider the various functions that vignettes must serve (e.g., Taylor, 2006; Wallander, 2012) and their relations to internal, external, and construct validity (Finger & Rand, 2005). First, a vignette must *simulate* certain aspects of real-world scenarios, often bearing some resemblance to situations encountered by the participant. This simulation function is a facet of *construct validity*, or the degree to which a variable approximates or measures the intended theoretical construct. Second, vignettes, and the differences between vignettes, are intended to *elicit* some kind of effect that is hypothesized to exist independently in the real world. This elicitation function relates to the study's *internal validity*, or the degree to which changes in the dependent variable can be accurately attributed to changes in the independent variable. Third, vignette studies should produce results that *generalize* to real-world situations encountered by the participants and others like them, reflecting the definition of *external validity*. These three components of validity are conceptually distinct but functionally interrelated (Finger & Rand, 2005). For example, poorly written, unrealistic vignettes (low construct validity) will likely fail to elicit responses from participants that can be clearly differentiated, thereby diminishing the integrity of causal inferences (low internal validity) and the generalizability of findings (low external validity).

Although these are important considerations for any methodology, vignette studies are unique in that so much of the weight of the study's validity is concentrated in one element: the vignettes. Due to the multipurpose nature of the vignettes used in experimental research, this methodology has considerable advantages as well as unique challenges and disadvantages. By way of comparison, traditional survey designs are recognized as having the potential for high external validity but numerous threats to internal validity. In contrast, traditional experiments are designed to ensure high internal validity, though this is often at the expense of external validity (Finger & Rand, 2005). Used appropriately, vignette studies have been recognized as a "hybrid" methodology that inherits the external validity strengths of survey research and the internal validity

strengths of experimental methods (Atzmüller & Steiner, 2010; Lauder, 2002; Ludwick & Zeller, 2001; Ludwick et al., 2004; Sniderman & Grob, 1996; Taylor, 2006).

Regarding the construct validity and simulation function of vignettes, it should be acknowledged that vignettes are *not* intended to re-create real-world situations. Rather, they are designed to approximate, isolate, manipulate, and measure key aspects of the decision-making processes that individuals use in real world situations (Alexander & Becker, 1978; Hughes, 1998; Veloski, Tai, Evans, & Nash, 2005; Wallander, 2012). Similarly, in regard to the external validity and generalizability of findings, participants' behavior in a vignette study is not intended to be interpreted as representative of their behavior in the real world, but rather as strong predictors or proxies for such behavior, given the circumstances approximated by the vignette (Hughes, 1998; Wallander, 2009, 2012). Thus, the more useful question is not whether the "vignette world" is equivalent to the "real world," but rather, whether the mental and behavioral processes used in the study are activated in a manner that is similar to the way in which these processes are used in real life.

It is notable that many of the criticisms about vignette artificiality are framed in arguments that cannot be definitively refuted. Indeed, one's hypothetical behavior in an experiment may differ from one's actual behavior in real life (Barter & Renold, 2000; Hughes, 1998; Jenkins, Bloor, Fischer, Berney, & Neale, 2010; Kim, 2012; Lauder, 2002; Ludwick & Zeller, 2001; Ludwick et al., 2004; Schoenberg & Ravdal, 2000). Further, most vignette studies do not (or, for ethical or logistical reasons, cannot) measure the correspondence between the hypothetical behavior and actual behavior. However, the available evidence addressing this issue suggests that participants respond to hypothetical and real-life scenarios in a similar manner. For example, in a follow-up study with nurses who had previously participated in a study using vignettes describing assaults on staff by patients, Lunza (1990) found that nurses' responses to vignette scenarios predicted their responses to real patient assault incidents that they subsequently encountered in their regular work. Similarly, family physicians made similar consultation referrals in response to case vignettes as to actual patients (Langley, Tritchler, Llewellyn-Thomas, & Till, 1991), and rheumatologists showed a nearly perfect correlation ($r = .90$) in their assessments of real patients and similar "paper patients" (Kirwan, De Santonge, Joyce, & Curry, 1983).

Several comparison studies and reviews have yielded similar conclusions. In a meta-analysis of 111 studies (Murphy, Herr, Lockhart, & Maguire, 1986), vignette methodologies demonstrated little difference from observations of actual behavior; although vignettes sometimes lead to larger effect sizes, which may reflect an increase in internal validity with some reduction in external validity. In studies with health care providers, clinical vignettes yield findings very similar to those using standardized patients (which researchers sometimes refer to as the "gold standard") for measuring physician care (Peabody et al., 2000, 2004; Shah, Edgar, & Evans, 2007; Veloski et al., 2005). In general, clinicians' perceptions of and responses to vignettes appear to resemble their responses to real life situations.

Practical and scientific utility of vignettes

In addition to the validity considerations discussed above, vignette methodologies offer several advantages in terms of their *utility* in the development and implementation of a research study. That is, in comparison to other methodologies, vignettes can be a useful tool for investigating a wide range of research subjects and question types (flexibility), with relatively low requirements in terms of time, personnel, funding, or other resources (efficiency). Using vignettes, researchers can represent particular scenarios accurately, concretely, and with a level of detail that supports their realism and credibility (Alexander & Becker, 1978; Atzmüller & Steiner, 2010; Wallander, 2009). The content of the vignette can be carefully tailored to provide a concise distillation of relevant information (comprised of experimental, controlled, and contextual content) while omitting unnecessary or irrelevant material. Vignette designs provide a feasible alternative to a range of other methodologies (e.g., record review, interview, focus group, diary, survey) and are well suited for almost any question about the judgments and decision-making processes of individuals in general, and of professionals in particular (Ganong & Coleman, 2006; Ludwick et al., 2004; Shah et al., 2007; Taylor, 2006; Wallander, 2009, 2012).

Indeed, research suggests that vignette methodologies are a valid, reliable, inexpensive, and practical method for assessing clinical practice (Bachmann et al., 2008; Veloski et al., 2005). For example, vignette studies assessing quality of care among outpatient physicians found vignettes to provide higher criterion validity, consistency, and variation in comparison to other research methodologies, including the much more expensive use of standardized patients (Peabody et al., 2000, 2004). Similar results have been found in cross-national studies, providing some evidence for the cross-cultural applicability and utility of clinical vignettes (Peabody & Liu, 2007; Spreng et al., 2014). Further, by using vignettes, these studies bypass a number of ethical concerns that can arise in studies with actual patients or confidential health data (Shah et al., 2007; Taylor, 2006), not to mention cost and feasibility constraints (Peabody et al., 2000; Veloski et al., 2005).

Vignettes can overcome several well-known internal validity weaknesses that accompany traditional survey research. A concrete, detailed, hypothetical vignette situation provides a better investigative vehicle than asking abstract questions about attitudes and perceptions (Alexander & Becker, 1978). In survey research, the use of pre-established questionnaires, checklists, and open-ended questions can introduce investigative bias, prompting effects, and an undesirable degree of specificity, ambiguity, or misinterpretation in participants' responses (Schoenberg & Ravdal, 2000). Vignettes can overcome these limitations by providing a consistent, non-personal frame of reference that allows participants to think beyond the idiosyncrasies of their own personal situation (Schoenberg & Ravdal, 2000).

The participant's "psychological distance" from the research question is critical in that it allows investigators to overcome various observer effects, including desirability or "yea-saying" effects (Alexander & Becker, 1978; Hughes & Huby, 2001; Schoenberg & Ravdal, 2000; Wason

Table 1 Recommendations for vignette content.

Vignettes should

1. Derive from the literature and/or clinical experience
2. Be clear, well-written, and carefully edited
3. Not be longer than necessary (typically between 50 and 500 words)
4. Follow a narrative, story-like progression
5. Follow a similar structure and style for all vignettes in the study
6. Use present tense (past tense only for history and background information)
7. Avoid placing the participant "in the vignette" (e.g., as first- or third-person character)
8. Balance gender and age across vignettes*
9. Be as neutral as possible with respect to cultural and socio-economic factors*
10. Resemble real people, not a personification of a list of symptoms or behaviors
11. Be relatable, relevant, and plausible to participants
12. Avoid "red herrings", misleading details, and bizarre content
13. Highlight the key variables of interest, facilitating experimental effects
14. Facilitate participant engagement and thinking by including vague or ambiguous elements
15. Cover all pertinent variables (or omit selected variables for specific purposes)

Key references: (Ganong & Coleman, 2006; Gould, 1996; Hughes, 1998; Hughes & Hughes & Huby, 2001; Jenkins et al., 2010; Kim, 2012; Veloski et al., 2005; Wallander, 2009).

* Exceptions may apply if these factors are included among the experimental variables.

et al., 2002; Wallander, 2009). By asking research questions and inferring conclusions indirectly, researchers can investigate sensitive or controversial issues and perspectives, attitudes, and biases that participants may not feel comfortable disclosing, or may even escape their own awareness (Alexander & Becker, 1978; Barter & Renold, 2000; Ganong & Coleman, 2006; Gould, 1996; Grønhøj & Bech-Larsen, 2010; Hughes, 1998; Hughes & Huby, 2001; Kim, 2012; Wallander, 2009). More importantly, researchers can determine how these biases influence participants' decisions and behaviors.

Any psychological research can be subject to the "Hawthorne effect," where participant behavior varies as a function of whether they think they are being observed. Gould (1996) argues that a major advantage of using vignettes is that participants are less likely to perceive and be influenced by the act of observation. Similarly, clinician participants are also subject to the "Sentinel effect," where their behavior may change as a result of being *evaluated* (or perceived as such) rather than simply observed (Veloski et al., 2005). Indeed, practitioners readily recognize their own motivation to justify their clinical decisions and give the "right" answer (Taylor, 2006), reflecting a natural propensity toward presenting oneself as a competent professional. Veloski et al. (2005) suggested that, scientifically, unidentified standardized patients may be the best method for overcoming these observer effects; ethically, however, this practice can be problematic. Considering the "distance" afforded by the vignette scenario, coupled with clear indications of confidentiality/anonymity and the non-evaluative nature of the study, vignette designs are perhaps the most effective means for minimizing these observer effects (Gould, 1996), while still allowing clinicians full knowledge of the nature of their participation (Veloski et al., 2005).

Regarding internal validity, vignette studies afford researchers control and standardization of vignette presentation, alongside systematic manipulation of key aspects of the vignette, thus allowing for rigorous causal inferences to

be drawn with respect to the unique and shared variance of multiple factors predicting clinician behavior (Alexander & Becker, 1978; Atzmüller & Steiner, 2010; Taylor, 2006; Veloski et al., 2005; Wallander, 2009). In addition, vignettes can be written in a way that overcomes the multicollinearity and nonorthogonality among variables as they occur in the real world, while still appearing just as realistic to participants (Ludwick et al., 2004; Taylor, 2006; Wallander, 2009). Vignette designs can facilitate increased analytic power and efficiency across a number of analytic methods, including regression, ANOVA, and multilevel modeling (Atzmüller & Steiner, 2010; Ludwick et al., 2004; Murphy et al., 1986; Wallander, 2009, 2012). Finally, clinician participants may find enjoyment, challenge, and creativity in participating in vignette studies (Schoenberg & Ravdal, 2000; Wason et al., 2002). Taylor (2006) argues that, to the extent that the design can increase participants' engagement (e.g., reading stories and solving problems rather than answering a list of survey questions), the data will be more accurate as a result.

Methodological recommendations

Based on the literature reviewed above, as well as our own research and clinical experience, we now offer several recommendations for the development and implementation of vignette studies examining clinicians' decision-making processes. Of course, specific prescriptions for study design would vary a great deal depending on the context and purpose of the study, so we focus only on general recommendations applicable to most studies. For clarity and brevity, we summarize our recommendations for vignette *content* in Table 1 and focus our discussion on procedures for vignette *development*.

As with any investigation, the research objectives should first be clearly defined and several methodological approaches should be considered. Vignettes are particularly

useful when the research question involves an understanding of causal associations related to phenomena that are inaccessible or unobservable for ethical, logistical, or validity reasons. A wide array of design variations (e.g., factorial survey, traditional experimental designs) can be considered and flexibly adapted to suit the purpose of the study. Factorial surveys lend themselves to investigations involving multiple continuous variables, whereas traditional experimental designs are often more effective for between-group or within-subject research questions.

Once a vignette-based design is selected, researchers should develop vignettes specifically to accomplish the study objectives with attention to internal, external, and construct validity. Researchers should initially develop a larger-than-necessary vignette population from which they can subsequently select or piece together the best vignettes to comprise the study sample (Atzmüller & Steiner, 2010). The final vignettes should be submitted to a panel of experts for further review and revision in order to improve the vignettes' clarity, cultural neutrality, and validity (Gould, 1996; Lauder, 2002; Veloski et al., 2005). The effectiveness of a review process can be increased by including multiple reviewers and providing specific instructions regarding the key components of the vignette that should be identified and rated (e.g., levels of functional impairment and symptom severity) and on what basis (e.g., clinical experience, comparison to a reference). As far as possible, the reviewers should be kept blind—for example, they may not need to know the precise methodological nature (experimental, contextual, or controlled) of the dimensions that they are asked to rate, or how specific vignettes fit into the larger experimental design. Next, after the vignettes are revised and finalized, they should be subjected to a separate pilot testing process (Flaskerud, 1979; Ganong & Coleman, 2006; Gould, 1996).

Empirical methods can be utilized to address the validity concerns discussed above. With respect to threats to internal validity, researchers can include items that serve as manipulation checks to determine the degree to which the vignette text elicited the desired effect (e.g., items assessing participants' perceptions of experimental aspects versus controlled aspects of the vignette). Regarding construct validity, items can be included to assess the degree to which the vignette resembled persons or scenarios encountered by the participant in real life. Following Lunza's (1990) example, external validity can be assessed by identifying the association between participants' hypothetical behavior and their real behavior in similar situations, possibly in a follow-up study.

Finally, to mitigate observer effects, clinicians should be told that the researchers' goal is to study the behaviors of a group, not to evaluate the behaviors of individuals (Veloski et al., 2005). Depending on the design of the study, researchers may also need to clarify a number of issues related to how participants are to respond—for example, as themselves, or as if they were a character in the vignette; describing what they *would* do or what they *should* do in that scenario; and acknowledging or ignoring real-life limitations (e.g., availability of specialists for referrals). Randomization and counterbalancing techniques should be used whenever possible and appropriate. Further, when designing studies where participants read multiple vignettes, researchers

should be mindful of whether this violates independence assumptions (Taylor, 2006) or is likely to induce carry-over effects (Sniderman & Grob, 1996). Caution should also be taken to ensure that participation is not excessively time-consuming, reading-intensive, boring, or burdensome (Ganong & Coleman, 2006; Kim, 2012; Wallander, 2009), and this is a particularly important concern in factorial designs and studies with multiple vignettes and item sets.

Example ICD-11 case-controlled field studies

WHO is currently developing the eleventh revision of the International Classification of Diseases and Related Health Problems (ICD-11), planned for approval by the World Health Assembly in 2017. The development and maintenance of the ICD is a core constitutional responsibility of WHO, a specialized agency of the United Nations whose mission and activities focus on global public health. The ICD is the international standard for health information to enable the assessment and monitoring of mortality, morbidity, and other relevant parameters related to health around the world. The WHO Department of Mental Health and Substance Abuse is responsible for coordinating the technical activities associated with the development of the ICD-11 chapter on Mental and Behavioural Disorders within the context of the overall revision framework (International Advisory Group for the Revision of ICD-10 Mental and Behavioural Disorders, 2011).

The WHO Department of Mental Health and Substance Abuse has undertaken a systematic program of global research, including *formative* field studies to inform the development of the proposed diagnostic guidelines for ICD-11 (Evans et al., 2013; Reed, Correia, Esparza, Saxena, & Maj, 2011; Reed et al., 2013; Roberts et al., 2012) and subsequent *evaluative* field studies to assess their validity, reliability, and clinical utility in practice around the world (International Advisory Group for the Revision of ICD-10 Mental and Behavioural Disorders, 2011; Reed, 2010).² A central aim of the ICD-11 development process is to improve the diagnostic classification system's *clinical utility*, which has been previously defined based on its value for communicating clinical information, its implementation characteristics in clinical settings (e.g., ease of use, accuracy, feasibility), and its usefulness for informing clinical decision making (Reed, 2010, p. 461). Along with advances in scientific knowledge, improving clinical utility has been identified by WHO as a central rationale for making changes (International Advisory Group for the Revision of ICD-10 Mental and Behavioural Disorders, 2011). Key questions for the current phase of field testing relate to whether the proposed ICD-11 diagnostic guidelines result in more consistent diagnoses than existing ICD-10 diagnostic guidelines and whether clinicians find them to be more clinically useful.

Based on the literature reviewed above and numerous discussions with collaborating researchers, it was determined that clinical vignettes would be a valid, feasible,

² Evaluative field studies are currently being conducted, with results to be disseminated in future publications. A forthcoming article (Keeley et al., 2015) provides an overview of the methodology and rationale of these studies.

and appropriate tool for assessing these questions. This was particularly the case given the goal of comparing diagnostic guidance developed for ICD-11 with existing diagnostic guidelines in ICD-10. In order to test the specific effect of the different guidelines, it would be necessary to control for the variability associated with clinical presentations to which the guidelines would be applied and to manipulate key variables of interest. As reviewed above, case-controlled studies using vignette methodologies are ideally and specifically suited to the examination of these research questions, and superior to more ecologically-based methods in which it would not be possible to control the case variability associated with real patients. In vignette-based studies, the characteristics of the cases to be rated can be standardized and systematically controlled or varied to include specific attributes that are relevant to the research question.

Thus, a series of case-controlled studies using vignettes, and disseminated via the internet, are being implemented to assess the utility of proposed changes to ICD-11 among a global, multilingual, and multidisciplinary sample of over 11,000 mental health professionals from more than 130 countries (see www.globalclinicalpractice.net). We provide a brief description of these studies here for illustration of the principles described above. Each of the ICD-11 case-controlled field studies is designed to assess the differences in clinicians' diagnostic behavior produced by the ICD-10 and the proposed ICD-11 diagnostic guidelines. These field studies generally follow a similar overarching design, with each study tailored to address questions specific to the content area.³ Clinician participants are first randomly assigned to use either ICD-10 or ICD-11 diagnostic guidelines, and they are asked to review the portions of this material relevant to the area of the study. Second, participants are presented with a series of two vignettes (counterbalanced), selected from a small pool of paired vignette material, carefully designed to investigate the effects of key changes proposed for ICD-11. Using the diagnostic guidelines presented, participants are asked to choose a diagnosis or indicate that no diagnosis is warranted. Because the study is electronically presented via the internet, subsequent questions can be adaptively programmed to facilitate a step-by-step re-evaluation using diagnostic questions, followed by items assessing various aspects of the vignette (e.g., severity, impairment), the guidelines (e.g., ease of use, goodness of fit, representativeness), and the clinicians' diagnostic response (e.g., confidence in diagnosis, differential diagnosis considerations). The design allows the researchers to draw causal inferences related to how differences between diagnostic systems (ICD-10 vs. ICD-11) and vignette material (first vs. second vignette) affect clinicians' diagnostic decision-making, including the accuracy, efficiency, and clarity of the diagnosis, and the clinical utility of the ICD-11 proposals.

In each study, the research team follows particular guidelines for the development, pre-testing, revising, and

pilot testing of the vignettes and the studies. For example, to investigate the proposed addition of Complex Post-Traumatic Stress Disorder to the ICD-11 diagnostic classification in the study for Disorders Specifically Related to Stress, vignettes were developed to reflect prototypic presentations of Post Traumatic-Stress Disorder (PTSD) and Complex PTSD, with different types of precipitating stressors and symptomatology. Vignettes were written by global members of the Working Group on Disorders Specifically Associated with Stress (Maercker et al., 2013) based on their clinical and research expertise in that area of psychopathology. General guidelines were provided parallel to the recommendations in Table 1. Specific instructions were also provided regarding the individual characteristics of each specific vignette based on the questions of interest for each paired comparison. Experimental elements (e.g., differences in type of trauma) were embedded within vignettes that were otherwise similar (i.e., controlled aspects). To address artificiality concerns, vignettes were to be written based on real cases that the Working Group members had seen in their clinical practice, following a familiar structure for the presentation of clinical information in case notes and including personal details (contextual content) that "fleshed out" the personhood of the case. Writers were asked to clearly differentiate present from historical information (e.g., through careful use of tenses in describing past traumatic events and current symptoms), and to avoid the inclusion of culturally specific details. Multiple drafts of each vignette were written so that the investigators select the vignettes (in some cases by revising or combining them) that best demonstrated the required characteristic given requirements of construct validity (e.g., the extent to which the vignette reflects a diagnosis of PTSD or Complex PTSD) and the intended experimental manipulation. Researchers then carefully edited the vignettes for narrativity, readability, consistency (e.g., length, presentation style), and cultural neutrality.

In each of these case-controlled field studies, the initial set of vignettes undergoes a rigorous pre-testing in which expert raters read each vignette, provide a diagnosis, confirm (or deny) the presence of key diagnostic features, indicating the specific passages of the vignette that support those conclusions, and complete ratings of the severity of the case. This pre-testing process was designed to evaluate whether expert raters reach consistent diagnostic conclusions. This provides a validation of the vignette content, so that variability in response patterns of participants can be confidently interpreted as reflecting changes in diagnostic reasoning rather than ambiguous content in the vignettes themselves. Where inconsistencies are noted by expert raters, the vignettes are clarified accordingly, or an alternative vignette is developed and pre-tested. Finally, the vignettes are embedded in the study itself and a different set of collaborators pilot tests the study to provide a final review of the survey programming, participant experience, and vignette material in context.

It is notable that the vignette developers and pre-testers in the ICD-11 vignette case-controlled field studies are always multidisciplinary and represent different world regions, including a high proportion who are from low- and middle-income countries and who are not native English speakers. These procedures help to ensure

³ This article focuses on ICD-11 case-controlled field studies using an experimental design comparing ICD-10 and ICD-11. Other ICD-11 case-controlled field studies use alternative research designs, for example based on Item Response Theory, that also make use of vignettes.

cultural neutrality of the symptoms and additional details of the vignette as well as reflecting the intended users of the ICD-11. After vignette studies are developed, tested and launched in English, they undergo similarly rigorous processes for translation, re-testing, and implementation in other languages, including Spanish, Chinese, French, Japanese, and Russian.

Conclusions

In this article, we have reviewed methodological issues and design variations related to using research methodologies utilizing vignettes, particularly as a way of investigating clinicians' decision-making behaviors. When used appropriately, vignettes can be a flexible, practical, and powerful tool for studying how various factors influence clinicians' judgments and decisions, allowing for a degree of experimental control over stimulus presentation that is typically not feasible or ethical using real patients. Within a sound research design, vignette studies can combine the strengths of survey and experimental methodologies, maximizing both internal and external validity. Although vignette methodologies are not without their limitations, they can be superior for investigating specific types of questions and highly generalizable to real world behavior among clinicians. Evidence supports the use of vignette studies to provide one perspective—often not accessible through alternative methods—into the mental and behavioral processes that clinicians use in their professional lives. Focusing on internal, external, and construct validity, we have provided several recommendations that can help improve the methodological rigor and scientific thinking regarding vignette-based research. The ICD-11 case-controlled field vignettes provide an example of well-designed vignette studies aimed at addressing specific research questions for which vignette methodologies are specifically suited. This is an extremely important use of vignette studies in the context of developing a global system of mental disorder classification and illustrates the potential for these methodologies in global, multidisciplinary, and multilingual research in health care.

References

- Alexander, C. S., & Becker, H. J. (1978). The use of vignettes in survey research. *Public Opinion Quarterly*, *42*, 93–104. <http://dx.doi.org/10.1086/268432>
- Atzmüller, C., & Steiner, P. M. (2010). Experimental vignette studies in survey research. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *6*, 128–138. <http://dx.doi.org/10.1027/1614-2241/a000014>
- Bachmann, L., Mühleisen, A., Bock, A., Riet, G., Held, U., & Kessels, A. (2008). Vignette studies of medical choice and judgment to study caregivers' medical decision behavior: systematic review. *BMC Research Methodology*, *8*. <http://dx.doi.org/10.1186/1471-2288-8-50>
- Barnhill, J. W. (Ed.). (2014). *DSM-5 clinical cases*. Washington, D.C.: American Psychiatric Publishing.
- Barter, C., & Renold, E. (2000). 'I wanna tell you a story': Exploring the application of vignettes in qualitative research with children and young people. *International Journal of Social Research Methodology*, *3*, 307–323. <http://dx.doi.org/10.1080/13645570050178594>
- Baudson, T. G., & Preckel, F. (2013). Teachers' implicit personality theories about the gifted: An experimental approach. *School Psychology Quarterly*, *28*, 37–46. <http://dx.doi.org/10.1037/spq0000011>
- Cullen, S. (2010). Survey-driven romanticism. *Review of Philosophy and Psychology*, *1*, 275–296. <http://dx.doi.org/10.1007/s13164-009-0016-1>
- Evans, S. C., Reed, G. M., Roberts, M. C., Esparza, P., Watts, A. D., Correia, J. M., Ritchie, P., Maj, M., & Saxena, S. (2013). Psychologists' perspectives on the diagnostic classification of mental disorders: Results from the WHO-IUPsyS Global Survey. *International Journal of Psychology*, *48*, 177–193. <http://dx.doi.org/10.1080/00207594.2013.804189>
- Faia, M. A. (1980). The vagaries of the vignette world: A comment on Alves and Rossi. *American Journal of Sociology*, *85*, 951–954.
- Finger, M. S., & Rand, K. L. (2005). Addressing validity concerns in clinical psychology research. In M. C. Roberts, & S. S. Iardi (Eds.), *Handbook of research methods in clinical psychology* (pp. 13–30). Malden, MA: Blackwell Publishing.
- Flaskerud, J. H. (1979). Use of vignettes to elicit responses toward broad concepts. *Nursing Research*, *28*, 210–212.
- Ganong, L. H., & Coleman, M. (2006). Multiple segment factorial vignette designs. *Journal of Marriage and Family*, *68*, 455–468. <http://dx.doi.org/10.1111/j.1741-3737.2006.00264.x>
- Gould, D. (1996). Using vignettes to collect data for nursing research studies: How valid are the findings? *Journal of Clinical Nursing*, *5*, 207–212. <http://dx.doi.org/10.1111/j.1365-2702.1996.tb00253.x>
- Grøndal, P., Grønnerød, C., & Sexton, J. (2009). A comparative case vignette study of decision making in forensic psychiatric cases. *International Journal of Forensic Mental Health*, *8*, 263–270. <http://dx.doi.org/10.1080/14999011003635639>
- Grønhoj, A., & Bech-Larsen, T. (2010). Using vignettes to study family consumption processes. *Psychology & Marketing*, *27*, 445–464. <http://dx.doi.org/10.1002/mar.20338>
- Herskovitz, M. J. (1950). The hypothetical situation: A technique of field research. *Southwestern Journal of Anthropology*, *6*, 32–40.
- Howie, P., Nash, L., Kurukulasuriya, N., & Bowman, A. (2012). Children's event reports: Factors affecting responses to repeated questions in vignette scenarios and event recall interviews. *British Journal of Developmental Psychology*, *30*, 550–568. <http://dx.doi.org/10.1111/j.2044-835X.2011.02064.x>
- Hughes, R. (1998). Considering the vignette technique and its application to a study of drug injecting and HIV risk and safer behaviour. *Sociology of Health & Illness*, *20*, 381–400. <http://dx.doi.org/10.1111/1467-9566.00107>
- Hughes, R., & Huby, M. (2001). The application of vignettes in social and nursing research. *Methodological Issues in Nursing Research*, *37*, 382–386. <http://dx.doi.org/10.1046/j.1365-2648.2002.02100.x>
- International Advisory Group for the Revision 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. <http://dx.doi.org/10.1002/j.2051-5545.2011.tb00022.x>
- Jenkins, N., Bloor, M., Fischer, J., Berney, L., & Neale, J. (2010). Putting it in context: The use of vignettes in qualitative interviewing. *Qualitative Research*, *10*, 175–198. <http://dx.doi.org/10.1177/1468794109356737>
- Jones, C., & Aronson, E. (1973). Attribution of fault to a rape victim as a function of respectability of the victim. *Journal of Personality and Social Psychology*, *26*, 415–419.
- Keeley, J. W., Reed, G. M., Roberts, M. C., Evans, S. C., Medina-Mora, M. E., Robles, R., Rebello, T., Gureje, O., First, M. B., Andrews, H. F., Ayuso-Mateos, J. L., Gaebel, W., Zielasek, J., & Saxena, S. (2015). *Developing a science of clinical utility in diagnostic classification systems: Field study strategies for*

- ICD-11 *Mental and Behavioural Disorders*. Manuscript submitted for publication.
- Kim, J. (2012). Scenarios in information seeking and information retrieval research: A methodological application and discussion. *Library & Information Science Research*, 34, 300–307. <http://dx.doi.org/10.1016/j.lisr.2012.04.002>
- Kirwan, J. R., De Saintonge, D. C., Joyce, C. R., & Currey, H. L. (1983). Clinical judgment in rheumatoid arthritis. I. Rheumatologists' opinions and the development of 'paper patients'. *Annals of the Rheumatic Diseases*, 42, 644–647.
- Langlely, G. R., Tritchler, D. L., Llewellyn-Thomas, H. A., & Till, J. E. (1991). Use of written cases to study factors associated with regional variations in referral rates. *Journal of Clinical Epidemiology*, 44, 391–402. [http://dx.doi.org/10.1016/0895-4356\(91\)90077-M](http://dx.doi.org/10.1016/0895-4356(91)90077-M)
- Lauder, W. (2002). Factorial survey methods: A valuable but under-utilised research method in nursing research? *Nursing Times Research*, 7, 35–43. <http://dx.doi.org/10.1177/136140960200700106>
- Levy Cole, K., Roberts, M. C., & McNeal, R. E. (1996). Children's perceptions of ill peers: Effects of disease, grade, and impact variables. *Children's Health Care*, 25, 107–115. <http://dx.doi.org/10.1207/s15326888chc2502.3>
- Ludwick, R., Wright, M. E., Zeller, R. A., Dowding, D. W., Lauder, W., & Winchell, J. (2004). An improved methodology for advancing nursing research: Factorial surveys. *Advances in Nursing Science*, 27, 224–238. <http://dx.doi.org/10.1097/00012272-200407000-00007>
- Ludwick, R., & Zeller, R. A. (2001). The factorial survey: An experimental method to replicate real world problems. *Nursing Research*, 50, 129–133. <http://dx.doi.org/10.1097/00006199-200103000-00009>
- Lunza, M. L. (1990). A methodological approach to enhance external validity in simulation based research. *Issues in Mental Health Nursing*, 11, 407–422.
- Macpherson, C. C., & Veatch, R. M. (2010). Medical student perceptions about bioethics. *Cambridge Quarterly of Healthcare Ethics*, 19, 488–496. <http://dx.doi.org/10.1017/S0963180110000381>
- Maercker, A., Brewin, C. R., Bryant, R. A., Cloitre, M., van Ommeren, M., Jones, L. M., Humayan, A., Kagee, A., Llosa, A. E., Rousseau, C., Somasundaram, D. J., Souza, R., Suzuki, Y., Weissbecker, I., Wessely, S. C., First, M. B., & Reed, G. M. (2013). Diagnosis and classification of disorders specifically associated with stress: Proposals for ICD-11. *World Psychiatry*, 12, 198–206. <http://dx.doi.org/10.1002/wps.20057>
- Maieron, M. J., Roberts, M. C., & Prentice-Dunn, S. (1996). Children's perceptions of peers with AIDS: Assessing the impact of contagion information, perceived similarity, and illness conceptualization. *Journal of Pediatric Psychology*, 21, 321–333.
- McCarron, M., & Stewart, D. (2011). A Canadian perspective on using vignettes to teach ethics in psychology. *Canadian Psychology*, 52, 185–191. <http://dx.doi.org/10.1037/a0024504>
- McGlynn, R. P., Megas, J. C., & Benson, D. H. (1976). Sex and race as factors affecting the attribution of insanity in a murder trial. *The Journal of Psychology*, 93, 93–99. <http://dx.doi.org/10.1080/00223980.1976.9921378>
- Murphy, K. R., Herr, B. M., Lockhart, M. C., & Maguire, E. (1986). Evaluating the performance of paper people. *Journal of Applied Psychology*, 71, 654–661.
- Peabody, J. W., & Liu, A. (2007). A cross-national comparison of the quality of clinical care using vignettes. *Health Policy and Planning*, 22, 294–302. <http://dx.doi.org/10.1093/heapol/czm020>
- Peabody, J. W., Luck, J., Glassman, P., Dresslhaus, T., & Lee, M. (2000). Comparison of vignettes, standardized patients, and chart abstraction: A prospective validation study of 3 methods for measuring quality. *Journal of the American Medical Association*, 283, 1715–1722. <http://dx.doi.org/10.1001/jama.283.13.1715>
- Peabody, J. W., Luck, J., Glassman, P., Jain, S., Hansen, J., Spell, M., & Lee, M. (2004). Measuring the quality of physician practice by using clinical vignettes: A prospective validation study. *Annals of Internal Medicine*, 141, 771–780. <http://dx.doi.org/10.7326/0003-4819-141-10-200411160-00008>
- Reed, G. M. (2010). Toward ICD-11: Improving the clinical utility of WHO's international classification of mental disorders. *Professional Psychology: Research and Practice*, 41, 457–464. <http://dx.doi.org/10.1037/a0021701>
- Reed, G. M., Correia, J. M., Esparza, P., Saxena, S., & Maj, M. (2011). The WPA-WHO global survey of psychiatrists' attitudes toward mental disorders classification. *World Psychiatry*, 10, 118–131. <http://dx.doi.org/10.1002/j.2051-5545.2011.tb00034.x>
- Reed, G. M., Roberts, M. C., Keeley, J., Hooppell, C., Matsumoto, C., Sharan, P., Robles, R., Carvalho, H., Wu, C., Gureje, O., Lea-Mateos, J. L., Flanigan, E. H., Corroia, J. M., Saxena, S., & Medina-Mora, M. E. (2013). Mental health professionals' natural taxonomies of mental disorders: Implications for the clinical utility of the ICD-11 and the DSM-5. *Journal of Clinical Psychology*, 69, 1191–1212. <http://dx.doi.org/10.1002/jclp.22031>
- Roberts, M. C., Johnson, A. Q., & Beidleman, W. B. (1984). The role of socioeconomic status on children's perceptions of medical and psychological disorders. *Journal of Clinical Child Psychology*, 13, 243–249. <http://dx.doi.org/10.1080/15374418409533197>
- Roberts, M. C., Reed, G. M., Medina-Mora, M. E., Keeley, J. W., Sharan, P., Johnson, D. K., Mari, J. de J., Ayuso-Mateos, J. L., Gureje, O., Xiao, Z., Maruta, T., Khoury, B., Robles, R., & Saxena, S. (2012). A global clinicians' map of mental disorders to improve ICD-11: Analysing meta-structure to enhance clinical utility. *International Review of Psychiatry*, 24, 578–590. <http://dx.doi.org/10.3109/09540261.2012.736368>
- Ross, S., Moffat, K., McConnachie, A., & Wilson, P. (1999). Sex and depression: A randomized vignette study of the utility of depression by general practitioners. *British Journal of General Practice*, 49, 17–21.
- Rossi, P. H., & Nock, S. L. (Eds.). (1982). *Measuring social judgments: The factorial survey approach*. Beverly Hills: Sage.
- Rossi, P. H., Sampson, W. A., Bose, C. E., Jasso, G., & Passel, J. (1974). Measuring household social standing. *Social Science Research*, 3, 169–190. [http://dx.doi.org/10.1016/0049-089X\(74\)90011-8](http://dx.doi.org/10.1016/0049-089X(74)90011-8)
- Schoenberg, N. E., & Ravdal, H. (2000). Using vignettes in awareness and attitudinal research. *International Journal of Social Research Methodology*, 3, 63–74. <http://dx.doi.org/10.1080/136455700294932>
- Shah, R., Edgar, D., & Evans, B. J. W. (2007). Measuring clinical practice. *Ophthalmic and Physiological Optics*, 27, 113–125. <http://dx.doi.org/10.1111/j.1475-1313.2006.00481.x>
- Sniderman, P. M., & Grob, D. B. (1996). Innovations in experimental design in attitude surveys. *Annual Review of Sociology*, 22, 377–399. <http://dx.doi.org/10.1146/annurev.soc.22.1.377>
- Spalding, N. J., & Phillips, T. (2007). Exploring the use of vignettes: From validity to trustworthiness. *Qualitative Health Research*, 17, 954–962. <http://dx.doi.org/10.1177/1049732307306187>
- Spreng, C. P., Ojo, I. P., Burger, N. E., Sood, N., Peabody, J. W., & Demaria, L. M. (2014). Does stewardship make a difference in the quality of care? Evidence from clinics and pharmacies in Kenya and Ghana. *International Journal for Quality of Health Care*. Advance online publication. <http://dx.doi.org/10.1093/intqhc/mzu054>
- Stokes, J., & Schmidt, G. (2011). Race, poverty and child protection decision making. *British Journal of Social Work*, 41, 1105–1121. <http://dx.doi.org/10.1093/bjsw/bcr009>

- Stolte, J. F. (1994). The context of satisficing in vignette research. *The Journal of Social Psychology, 134*, 727–733. <http://dx.doi.org/10.1080/00224545.1994.9923007>
- Taylor, B. J. (2006). Factorial surveys: Using vignettes to study professional judgment. *British Journal of Social Work, 36*, 1187–1207. <http://dx.doi.org/10.1093/bjsw/bch345>
- Veloski, J., Tai, S., Evans, A. S., & Nash, D. B. (2005). Clinical vignette-based surveys: A tool for assessing physician practice variation. *American Journal of Medical Quality, 20*, 151–157. <http://dx.doi.org/10.1177/1062860605274520>
- Wainwright, P., Gallagher, A., Tompsett, H., & Atkins, C. (2010). The use of vignettes within a Delphi exercise: A useful approach in empirical ethics. *Journal of Medical Ethics, 36*, 656–660. <http://dx.doi.org/10.1136/jme.2010.036616>
- Wallander, L. (2009). 25 years of factorial surveys in sociology: A review. *Social Science Research, 38*, 505–520. <http://dx.doi.org/10.1016/j.ssresearch.2009.03.004>
- Wallander, L. (2012). Measuring social workers' judgments: Why and how to use the factorial survey approach in the study of professional judgments. *Journal of Social Work, 12*, 364–384. <http://dx.doi.org/10.1177/1468017310387463>
- Walster, E. (1966). Assignment of responsibility for an automobile accident. *Journal of Personality and Social Psychology, 3*, 73–79.
- Wason, K. D., Polonsky, M. J., & Hyman, M. R. (2002). Designing vignette studies in marketing. *Australasian Marketing Journal, 10*, 41–58. [http://dx.doi.org/10.1016/S1441-3582\(02\)70157-2](http://dx.doi.org/10.1016/S1441-3582(02)70157-2)
- Williamson, J. W. (1965). Assessing clinical judgment. *Journal of Medical Evaluation, 40*, 180–187.
- World Health Organization (2014). Classifications: Revision of the International Classification of Disease (ICD). Retrieved from <http://www.who.int/classifications/icd/ICDRevision>.