Improving the diagnostic criteria for alcohol use disorders through survey methodology and cognitive interviewing

LOUISE MEWTON,¹ TIM SLADE,¹ MAREE TEESSON,¹ SONJA MEMEDOVIC¹ & ROBERT F. KRUEGER²

1 National Drug & Alcohol Research Centre, University of New South Wales, Sydney, Australia

2 Department of Psychology, University of Minnesota, Minneapolis, MN, USA

Key words

Alcohol, diagnosis, DSM

Correspondence

Louise Mewton, Clinical
Research Unit for Anxiety and
Depression, 394–404 Victoria
Street, Darlinghurst, NSW 2010,
Australia.

Telephone (+61) 2 8382 1437

Fax (+61) 2 8382 1404

Email: louisem@unsw.edu.au

Received 13 September 2012;

revised 12 February 2013;

accepted 30 May 2013

Abstract

This study aims to identify problems in the structure and wording of questions designed to operationalize four DSM-IV diagnostic criteria for alcohol use disorders (AUDs): (1) use of alcohol in hazardous situations (*hazard*); (2) *tolerance*; (3) use of alcohol in larger amounts/longer periods than intended (*larger/longer*); (4) unsuccessful attempts to cut down or control alcohol use (*quit/cut down*). Ten experts appraised the questions related to these criteria in the WMH-CIDI according to a standardized checklist. These experts identified three main problems: (1) the double-barrelled nature of some of the questions; (2) definitional issues; and (3) unclear thresholds for criterion endorsement. Cognitive interviews of 100 young adult drinkers aged 18–24 were then conducted. The double-barrelled nature of the DSM-IV criteria led to their subsequent over- or under-endorsement. Key terms in the questions under investigation were defined inconsistently. There was also a large amount of variability in the thresholds at which *larger/longer* and *quit/cut down* were endorsed. Many of these problems could be linked back to the DSM-IV text. The findings raise questions as to the validity of AUD diagnoses when established via structured diagnostic interview. Further research should focus on testing alternative structure and wording of key AUD criteria to ensure accurate operationalizations of these criteria in structured diagnostic interviews. Copyright © 2014 John Wiley & Sons, Ltd.

Introduction

Mental health surveys of the general population aim to obtain reliable and valid information regarding the prevalence and correlates of mental disorders, as well as treatment history and the burden of mental disorders on both the individual and society (Kessler and Ustun, 2004). These large-scale, population based surveys often depend on lay administered structured diagnostic

interviews that operationalize diagnostic criteria for psychiatric diagnoses as literally as possible. Whilst diagnosis in practice involves clinician judgment, structured interviews diagnose mental illness according to carefully constructed diagnostic algorithms that are based on strict translations of diagnostic criteria. The development of these interviews has led to the proliferation of large-scale surveys that provide important information regarding the epidemiology of mental disorders in the general

population. Given the impact of epidemiological research on public policy and resource distribution, it is critical that these structured interviews operationalize diagnostic criteria in a manner that is consistent with the clinical phenomena they represent.

Recently we investigated four diagnostic criteria for alcohol abuse and dependence as operationalized in the World Mental Health Composite International Diagnostic Interview (WMH-CIDI), a structured diagnostic interview that has been used to estimate the prevalence and correlates of the major mental disorders internationally (Kessler and Ustun, 2004). These criteria (listed in Table 1) have previously been identified as particularly problematic when applied to young adults in the general population. For example, core alcohol dependence symptoms such as *tolerance* and impaired control over alcohol use (*larger/*

longer and *quit/cut down*) are endorsed at very high rates amongst younger populations (Chung *et al.*, 2001; Chung and Martin, 2002; Mewton *et al.*, 2011b, 2011c). Psychometric research has also shown that these criteria demonstrate age-based bias (Martin *et al.*, 2006; Mewton *et al.*, 2011b, 2011c). Meanwhile criteria related to alcohol abuse, such as alcohol consumption in hazardous situations (*hazard*), appear normative in younger populations and do not demonstrate a strong relationship with underlying alcohol use disorder (AUD) psychopathology (Mewton *et al.*, 2011a, 2011b, 2011c). Using survey methodology and cognitive interviewing techniques, we found that the underlying intent and clinical significance of these criteria was not always successfully captured in structured diagnostic interviews (Slade *et al.*, 2012). Young adults from the general population were shown to endorse

Table 1. DSM-IV wording and WMH-CIDI operationalizations of *hazard*, *tolerance*, *larger/longer* and *quit/cut down*

Criterion	DSM-IV wording	WMH-CIDI item wording	Percentage endorsing (WMH-CIDI)	Percentage endorsing (cognitive interview)
<i>Hazard</i>	Recurrent substance use in situations in which it is physically hazardous (e.g. driving an automobile or operating a machine when impaired by substance use)	"Were there times when you were often under the influence of alcohol in situations where you could get hurt, for example when riding a bicycle, driving, operating a machine, or anything else?"	24%	-
<i>Tolerance</i>	A need for markedly increased amounts of alcohol to achieve intoxication or desired effect, or markedly diminished effect with continued use of the same amount of alcohol	"Did you need to drink a larger amount of alcohol to get an effect, or did you find that you could no longer get a 'buzz' or a high on the amount you used to drink?"	45%	43%
<i>Larger/longer</i>	Alcohol is often taken in larger amounts or over a longer period than was intended.	A. "Did you have times when you started drinking even though you promised yourself you wouldn't, or when you drank a lot more than you intended?" B. "Were there times when you drank more frequently or for more days in a row than intended?" C. "Did you have times when you started drinking and became drunk when you didn't want to?"	93%	85%
<i>Quit/cut down</i>	A persistent desire or unsuccessful efforts to cut down or control alcohol use	"Were there times when you tried to stop or cut down on your drinking and found that you were not able to do so?"	21%	34%

criteria related to impaired control over alcohol use (i.e. *quit/cut down* and *larger/longer*) at very high rates. These criteria were designed to address compulsive use of alcohol that is central to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) diagnosis of alcohol dependence (Edwards and Gross, 1976). However, the cognitive interviews revealed that the majority of young adults endorsed these criteria on the basis of social, rather than compulsion-based, reasons (Slade *et al.*, 2012). These findings are similar to those reported in previous cognitive interviewing studies focusing on the DSM-IV AUD criteria (Caetano, 1999; Chung and Martin, 2005). We concluded that the intent of these criteria was not explicit in the DSM-IV text and subsequent operationalizations of these criteria, which are often literal translations of this text, failed to capture their clinical significance (Slade *et al.*, 2012). This resulted in the over-endorsement of these criteria which led to a considerable inflation of the rates of alcohol dependence amongst young adults in this sample.

The current paper reports further findings from this study, focusing specifically on the structure and wording of these criteria in structured diagnostic interviews. Previous research has shown that minor changes to the structure and wording of diagnostic interview questions may lead to large discrepancies in important outcomes, such as overall prevalence of disorder. For example, minor language differences in the operationalization of a single DSM-IV AUD criterion (*hazard*) have resulted in considerable discrepancies in the prevalence of AUDs across two national surveys of the US population (Vergés *et al.*, 2011), as well as differences in prevalence rates between the United States and Australia (Agrawal *et al.*, 2011; Mewton *et al.*, 2011a). Similarly, the prevalence of adolescent AUDs has been shown to differ considerably depending on how the *tolerance* and *larger/longer* criteria have been worded (Chung *et al.*, 2002; Harford *et al.*, 2009). Inconsistent operationalizations of the AUD criteria therefore have a significant impact on the estimated prevalence of these disorders.

Through the novel use of survey methodology and cognitive interviewing, the current paper aims to identify problems in the structure and wording of questions designed to operationalize four DSM-IV diagnostic criteria for AUDs. Whilst the findings are likely to have implications for all age groups, the current study will focus on the criteria particularly as applied to young adults (aged 18–24). This research is necessary to ensure consistent measurement and accurate prevalence estimates of AUDs in the general population.

Methods

The current study focused on the operationalization of the DSM-IV AUD criteria in the WMH-CIDI (Kessler and Ustun, 2004). Table 1 provides a description of the four target criteria as described in the DSM-IV text, as well as the wording of the survey items that operationalize each criterion in the WMH-CIDI. Consistent with previous research (Chung and Martin, 2005), two separate and sequential methods were used to identify potential problems in the structure and wording of the WMH-CIDI operationalization of AUD criteria: the Question Appraisal System (QAS) and individual cognitive interviews. The QAS is a structured tool designed to provide a preliminary systematic appraisal of potential problems in the content, structure and wording of survey items (Willis and Lessler, 1999). The issues identified during the QAS phase were then translated into question probes to be implemented in individual cognitive interviews. These methods have been described in more detail previously (Slade *et al.*, 2012).

Question Appraisal System (QAS)

The QAS, developed by the Research Triangle Institute, is a standardized checklist of criteria against which questionnaire items are appraised (Willis and Lessler, 1999). The QAS is based on several previous question appraisal systems and focuses on question characteristics that are likely to present problems when administered in the field. The development of the QAS has been described in more detail elsewhere (Willis and Lessler, 1999). The sample recruited for the QAS phase of the current study consisted of 10 experts in the collective fields of survey development, AUD and classification research. Experts were recruited from within the DSM-5 substance use disorders work group as they represent psychiatrists and psychologists directly responsible for revision of the diagnostic criteria for the substance use disorders. Researchers and clinicians with a known interest in the classification of AUDs in young adults were also approached to participate in the study. Training of the experts involved manual-based instruction, which is sufficient for the appropriate and effective use of the QAS (Willis and Lessler, 1999). Experts were asked to read and study each of the WMH-CIDI questionnaire items listed in Table 1 and note down any problems against eight standardized criteria. These criteria seek to identify problems: (1) with the readability of the item; (2) with introductions, instructions or explanations from the respondent's perspective; (3) with communicating the intent or meaning to the respondent; (4) with embedded assumptions; (5) as the result of limitations of

memory; (6) associated with querying topics that may be sensitive to the respondent; (7) with the adequacy of the response range provided; (8) any problems not already identified earlier.

Cognitive interview

Data from the QAS were synthesized and used to guide the construction of a series of structured and standardized verbal probes, each of which was designed to address an assumption or source of misinterpretation identified by the experts during the QAS phase. These verbal probes formed the content of the cognitive interviews. The sample recruited for the individual cognitive interviews consisted of 100 18–24 year old community volunteers (50% male, 73% Australian born) recruited through media and advertisement campaigns. The participants were evenly spread across the 18–24 year old age range (mean age = 20.6 years). For most of the sample, the highest qualification obtained was completion of high school (74%, $n = 74$). Furthermore, 24% ($n = 24$) were undertaking or had completed trade or university qualifications, whilst 2% ($n = 2$) had not completed high school. To be eligible for the study, participants needed to meet the standard screening criteria for the WMH-CIDI AUD module (i.e. consumption of alcohol on at least three days per week over the past 12 months or, if alcohol was consumed less frequently, to have consumed at least three standard drinks on days when they drank). Our interviewed sample had relatively heavy past year drinking histories. The majority (59%) drank 1–2 days per week and on days when they drank they drank an average of 6.5 standard drinks (standard deviation = 3.5). More detailed characteristics of the sample have been reported previously (Slade *et al.*, 2012). Participants were reimbursed \$50 for their time and travel expenses.

The cognitive interviews were conducted at the National Drug and Alcohol Research Centre, University of New South Wales, Sydney. Informed consent was obtained prior to each interview. All forms of cognitive interviewing involve the administration of questionnaire items to a participant while collecting additional information relevant to their responses. One method of cognitive interviewing involves the “think aloud” procedure in which interviewers encourage respondents to verbalize their thoughts while answering questions (Willis, 2005). This process is completed with minimal input from the interviewer and can be concurrent (as the respondent answers the question), or retrospective (immediately after the respondent answers the question; Beatty and Willis, 2007). A second cognitive interviewing technique involves

the administration of direct “verbal probes” designed to explore the survey response process. As with the “think aloud” procedure, these probes can be administered concurrently or retrospectively (Willis, 2005). These specific probes are developed by the researcher prior to the cognitive interview and are often based on the results of systematic expert reviews designed to identify potential problems in the questionnaire items, such as the QAS (Beatty and Willis, 2007). The “verbal probing” method involves considerably more input from the interviewer, allowing the researcher to test specific hypotheses, as well as control the structure and content of the interview (Willis, 2005). The goals of the researcher often determine which interviewing procedure is adopted (i.e. “think aloud” or “verbal probing”), as well as the timing of the cognitive interview (i.e. concurrent or retrospective).

Retrospective verbal probing techniques were used in the current study in order to address the specific hypotheses derived from the QAS phase. The first part of the interview consisted of diagnostic assessment according to the AUD module of the WMH-CIDI. The WMH-CIDI was administered as a means of establishing criterion endorsement rates, as well as the rates of DSM-IV alcohol abuse and dependence in the current sample. Standard WMH-CIDI sequencing meant that if the respondent endorsed one of the *larger/longer* questions, they were not asked the subsequent questions related to the *larger/longer* criterion. The cognitive interviews commenced immediately following the WMH-CIDI assessment. During the cognitive interviewing phase, respondents were asked questions related to all four criteria regardless of whether they endorsed them in the WMH-CIDI. The cognitive interviews were tape recorded and transcribed by an independent transcription service. Answers to the cognitive interview questions were summarized across the total sample according to a comprehensive predetermined coding scheme. Of this data 25% was re-coded by an independent research associate and the interrater reliability was 0.84 (Cohen’s kappa). The data were analysed in IBM SPSS Statistics v19.0. This study received ethics approval (reference no.: HREC10032) from the Ethics Secretariat of the University of New South Wales.

Results

Question Appraisal System (QAS)

Based on responses to the QAS, the experts identified three major issues with regards to the WMH-CIDI questions:

- (1) Double-barrelled questions. The WMH-CIDI questions for *tolerance*, *larger/longer* and *quit/cut down*

were identified as double-barrelled in nature. A question is considered double-barrelled in nature when two queries are present within a single probe. Given that a single answer is requested, it is unclear as to which component part of the criterion the respondent is actually endorsing. The double-barrelled nature of the question may also cause confusion in the respondent and should generally be avoided in question design (Kelley *et al.*, 2003).

- (2) Definitional issues. The experts also noted that specific terms in each of the WMH-CIDI questions were ill-defined and open to multiple interpretations by the respondents.
- (3) Thresholds for criterion endorsement. Finally, the experts also identified that the thresholds above which a criterion is considered present is not explicit within the WMH-CIDI questions. The behaviours represented by the criteria may not be considered pathological in and of themselves. It is the intensity or frequency with which these behaviours are experienced that suggests a substance use disorder (Caetano, 1999). The experts indicated that thresholds for criterion endorsement needed to be explicit for the *larger/longer* and *quit/cut down* criteria.

The problems related to each of the four target AUD criteria are discussed in more detail later. The source of these problems in the DSM-IV text is also discussed.

Hazard

Double-barrelled question: Not reported by the experts.

Definitional issues: The experts identified problems with the *hazard* query that were mainly definitional in nature. For example, “times” could be referring to a period of time, or it could be referring to specific occasions or instances. With reference to the intent of the DSM-IV criteria, it is apparent that “times” in the *hazard* probe is intended to mean periods of time, as indicated by the phrase “were there times when you were often ...”. The inclusion of “times” in this probe reflects the fact that the respondent is unlikely to have engaged in these behaviours frequently (or “often”) over the past 12 months, but there may have been periods of time (i.e. holidays) when these behaviours may have occurred more frequently. However, the resulting inclusion of the word “often” may be confusing for the respondent. For the *hazard* probe, the experts also noted that the term “under the influence” required more clarification. This term was used to operationalize the ambiguous DSM-IV term “substance use” which is open to interpretation in the DSM-IV text.

Thresholds for endorsement: Not reported by the experts.

Tolerance

Double-barrelled question: The WMH-CIDI query for *tolerance* was identified as double-barrelled in nature. The respondent was required to provide a single answer to a question which asked (1) whether they “needed to drink a larger amount of alcohol to get an effect” and (2) whether they “could no longer get a ‘buzz’ or a high on the amount [they] used to drink”.

Definitional issues: For the *tolerance* probe, the term “larger”, which operationalizes the ill-defined “marked increase” referred to in DSM-IV, could refer to an increase in the number of drinks consumed, or a change in the strength of alcohol beverage consumed (i.e. moving from beer which has a low alcohol content to spirits which have a higher alcohol content). “Effect”, which refers to the “intoxication or desired effect” in the DSM text, was also identified as open to interpretation. “The amount you used to drink” was considered ill-defined, implying a reference period which is not made explicit in the DSM or WMH-CIDI. Finally, the experts indicated that the WMH-CIDI terms “buzz” and “high”, used to operationalize the ambiguous term “effect” in the DSM text, may be more relevant to illicit substance use rather than the effects of alcohol.

Thresholds for endorsement: Not reported by the experts.

Larger/longer

Double-barrelled question: Parts A and B of the *larger/longer* WMH-CIDI operationalization were identified as double-barrelled in nature. Both questions require a single answer to two separate questions.

Definitional issues: As with question for the *hazard* criterion, each question operationalizing the *larger/longer* criterion included the term “times” and it is unclear what this term is referring to.

Thresholds for endorsement: The threshold above which the *larger/longer* criterion would be considered present is not explicit in each of the WMH-CIDI questions targeting this criterion. In the DSM-IV text, the *larger/longer* criterion specifies that the substance is “often” taken in larger amounts or over a longer period than is intended. It appears that the term “times” in each question addressing the *larger/longer* criterion is intended to operationalize the frequency with which these behaviours occur.

However, the meaning of the term “often” is not explicit in the DSM-IV criteria set and, with the absence of specific thresholds in the DSM, this criterion has been subsequently ill-defined in diagnostic instruments.

Quit/cut down

Double-barrelled question: The experts identified the WMH-CIDI query for *quit/cut down* as double-barrelled in nature. The respondents are required to provide a single answer to two implicit questions: (1) Were you unable to stop drinking?; and (2) Were you unable to cut down on drinking?

Definitional issues: The question for *quit/cut down* in the WMH-CIDI also included the term “times” and it is unclear what this term is referring to. For *quit/cut down*, the experts also identified problems with the terms “stop” and “cut down”. Whilst “stop” and “cut down” are intended to refer to a serious long-term attempt to quit or reduce alcohol consumption, this intention is not explicit in the WMH-CIDI query or the DSM text, and may refer to “stop” or “cut down” during a single drinking occasion.

Thresholds for endorsement: In order to establish a threshold for endorsement, the DSM-IV text states that there needs to be a “persistent” desire or unsuccessful attempts to cut down or control substance use. As with *larger/longer*, this criterion appears to be operationalized in the WMH-CIDI through the term “times” which is open to respondent interpretation. In the absence of a clear definition of “persistent” in the DSM-IV text, the subsequent operationalization of this term is ambiguous.

Cognitive interviews

Following the administration of the QAS, the cognitive interviewing script was constructed. Each question in the cognitive interviewing script was designed to address each of the problems identified by the experts during the QAS phase. For each criterion, the respondents were first asked whether they had any trouble responding to the question during the diagnostic assessment. This was to determine any problems with the WMH-CIDI queries from the respondent’s perspective. In order to determine which aspect of the WMH-CIDI probes respondents were endorsing, double-barrelled questions (*tolerance*, *larger/longer* and *quit/cut down*) were then broken down into their component parts and asked separately for each respondent. To address the experts’ concerns about definitional issues, respondents were then asked to define

terms identified as ambiguous by the experts in the QAS (all four target criteria). Table 2 presents the definitional issues identified by the experts and the cognitive interview questions designed to address these issues. Finally, respondents were asked questions designed to establish quantifiable thresholds for endorsement of the *larger/longer* and *quit/cut down* criteria. Table 3 presents the problems identified by the experts in relation to these two criteria, as well as the cognitive interview questions designed to address these problems.

The cognitive interviewing results for each of the WMH-CIDI questions relating to the four target criteria are discussed in more detail later.

Hazard

Problems from the respondents perspective: Twenty-six per cent ($n = 26$) of the sample indicated that they had problems with the *hazard* probe. Twenty-three of these 26 people had problems related to the communication of the meaning or intent of the question (i.e. clarity), one person indicated problems with the underlying assumptions made by the question and two people indicated that they may not know or may have trouble remembering the information.

Double-barrelled question: Not included in the cognitive interview.

Definitional issues: For the *hazard* criterion, 82% ($n = 82$) defined the word “times” in terms of single occasions/instances, rather than a period of time. The respondent definitions of other ambiguous terms in the *hazard* item are listed in Table 2.

Thresholds for endorsement: Not included in the cognitive interview.

Tolerance

Problems from the respondents perspective: Fourteen per cent ($n = 14$) had problems with the *tolerance* probe. Eleven of these 14 people had a problem with the clarity of the question, one person identified underlying assumptions as the main problem and one person indicated that a lack of knowledge/memory was the main problem.

Double-barrelled question: Fifty per cent of the sample answered positively to the first part of the *tolerance* WMH-CIDI probe which queried whether they had needed to drink a larger amount to get an effect. When asked the second part of the *tolerance* WMH-CIDI probe (whether they

Table 2. Definitional issues identified in the QAS, verbal probes addressing each issue, and participant response codes

Criterion	Problem identified in QAS	Verbal probe in cognitive interview	Participant response codes	Per cent
<i>Hazard</i>	"Times" is not defined for the respondent. Does it mean occasions or instances of periods of time?	When I asked you "were there times", what did you take "times" in this question to mean?	Periods of time	18%
			Occasions/instances	82%
	"Often" is not defined for respondents	When I asked whether you were "often" under the influence of alcohol in situations where you could get hurt, what did you consider the word "often" to mean?	Quantity (i.e. "once")	8%
			Frequency (i.e. "weekly")	9%
			Quantity-frequency (i.e. "twice a month")	11%
			Re-defined correctly	43%
			Re-defined incorrectly	9%
	"Under the influence" is not defined for respondents; does it mean drunk, tipsy or is it referring to legal driving limits (i.e. BAC < 0.05)	And when I said "under the influence", what did you take this to mean?	Drunk	46%
			Tipsy	9%
			Any alcohol use	12%
	<i>Tolerance</i>	When I asked you whether you ever drank a "larger" amount of alcohol to get an effect, what did you consider the word "larger" to mean in this question?	Legal limits (i.e. BAC > 0.05)	7%
			Strength (i.e. "moving from beer to spirits")	4%
	"Larger" is not defined for the respondent; does it mean quantity or strength of alcoholic beverage (i.e. moving from beer to spirits)?	Reference period (i.e. drinking more than I used to a year ago)	Reference period	53%
	"Effect" is not defined for the respondent; does it mean drunk, tipsy or any alcohol?	And when I asked you whether you ever drank a larger amount of alcohol to get an "effect" what did you think the word "effect" refers to in this question?	Drunk	59%
			Tipsy	12%
	"The amount you used to drink" implies a reference period which is not specified clearly	When answering this question, what did you consider to be "the amount you used to drink"?	Reference period (i.e. "the number of drinks I had a year ago")	29%
			Reference period without quantity (i.e. "one year ago")	7%
	"Buzz" is not defined and may be more associated with drugs rather than alcohol	And what does the word "buzz" mean to you?	Illicit substances	4%
			Drunk	26%
<i>Larger/longer (A)</i>	"High" is not defined and may be more associated with drugs rather than alcohol	And what does the word "high" mean to you?	Tipsy	23%
			Illicit substances	13%
	"Times" is not defined for the respondent. Does it mean occasions or instances of periods of time?	When I asked you "were there times", what did you take "times" in this question to mean?	Drunk	22%
			Tipsy	3%
			Periods of time	23%
			Occasions/instances	77%

(Continues)

Table 2. (Continued)

Criterion	Problem identified in QAS	Verbal probe in cognitive interview	Participant response codes	Per cent
<i>Larger/longer (B)</i>	"Times" is not defined for the respondent. Does it mean occasions or instances of periods of time?	When I asked you "were there times", what did you take "times" in this question to mean?	Periods of time Occasions/instances	38% ($n=8/21$) 62% ($n=13/21$)
<i>Larger/longer (C)</i>	"Times" is not defined for the respondent. Does it mean occasions or instances of periods of time?	When I asked you "were there times", what did you take "times" in this question to mean?	Periods of time Occasions/instances	37% ($n=3/8$) 63% ($n=5/8$)
<i>Quit/cut down</i>	"Times" is not defined for the respondent. Does it mean occasions or instances of periods of time?	When I asked you "were there times", what did you take "times" in this question to mean?	Periods of time Occasions/instances	44% 56%
	"Stop" is not defined for the respondent; is it a serious attempt to stop drinking, or stop drinking after the respondent has started drinking in a single drinking episode?	What did you take the word "stop" to mean in this question?	Period of time Single occasion	33% 15%
	"Cut down" is not defined for the respondent; does it mean cut down on quantity or frequency? Is it for a long period of time or just for a single drinking episode?	What did you take the phrase "cut down" to mean in this question?	Quantity	68%
			Frequency Quantity-frequency	4% 4%

Note: BAC, blood-alcohol concentration.

could no longer get a "buzz" or "high" on the amount they used to drink), 40% ($n=40$) of the sample responded positively. Table 1 presents the percentage of the sample endorsing the criterion initially during WMH-CIDI assessment, as well as the percentage of people endorsing *tolerance* during the cognitive interview when the two inherent questions were asked separately. Two people who had endorsed the *tolerance* criterion in the WMH-CIDI did not endorse this criterion when the constituent parts of this criterion were asked separately in the cognitive interview.

Definitional issues: The respondent definitions of ambiguous terminology in the *tolerance* item are listed in Table 2.

Thresholds for endorsement: Not included in the cognitive interview.

Larger/longer

Problems from the respondents perspective: Fifteen per cent ($n=15$) had problems with *larger/longer (A)*, nine of which identified problems with clarity and a further six identifying problems with the underlying assumptions of the question. Of the 21 people queried about *larger/longer (B)*, three respondents indicated they had problems with the clarity of this question, whilst none of the eight people queried about *larger/longer (C)* indicated problems with this question.

Double-barrelled question: When asked to focus on the first part of *larger/longer (A)*, 63% ($n=63$) of the respondents indicated they had drunk despite promising themselves that they would not drink. When asked to focus on the second part of *larger/longer (A)*, 63% also indicated that they had drunk more than they had intended. For *larger/*

Table 3. Questions aimed at establishing thresholds for criterion endorsement and average number of times respondents reported the relevant behaviours related to *larger/longer* and *quit/cut down*

Criterion	Verbal probe in cognitive interview	N ^a	Average number of times behaviour reported (SD)	Range
<i>Larger/longer (A)</i>	In the past 12 months, how many times have you drunk even though you promised yourself you wouldn't?	63	5.3 (SD = 3.6)	1–50
	On these times when you drank even though you promised yourself you wouldn't, how many standard drinks did you tend to have?	63	5.5 (SD = 3.2)	1–15
	How many standard drinks were you generally drinking before you put these limits on your drinking?	75	7.9 (SD = 4.1)	1–208
	How many standard drinks did you want to limit yourself to?	75	4.0 (SD = 2.1)	2–20
	<i>Average reduction in number of drinks</i>	75	4.0 (SD = 3.6)	0–18
	How many times have you drunk more than X standard drinks when you didn't intend to?	63	7.8 (SD = 13.7)	0–10
	On those times when you drank more than X standard drinks when you didn't intend to, how many standard drinks did you tend to have?	63	4.6 (SD = 4.7)	0–31.5
<i>Larger/longer (B)</i>	In the past 12 months, how many times did you drink more than X although you didn't intend to?	5	3.8 (SD = 2.4)	2–8
	In the past 12 months, how many times did you drink for more than X days in a row when you didn't intend to?	5	4.2 (SD = 2.4)	2–8
	In the last 12 months, when you intended to drink for X days in a row, but ended up drinking more frequently, how many days in a row did you end up drinking?	5	5 (SD = 4.1)	2–12
	<i>Average number of additional days drinking</i>	5	3.2 (SD = 2.8)	1–8
<i>Quit/cut down</i>	In the past 12 months, how many times did you try to stop drinking and found you couldn't do so?	21	3.8 (SD = 3.4)	1–13
	How many standard drinks were you drinking when you decided to cut down?	53	7.6 (SD = 3.3)	2–15
	How many standard drinks did you want to limit yourself to when you cut down?	53	3.0 (SD = 2.3)	0–10
	<i>Average reduction in number of drinks</i>	53	4.8 (SD = 3.2)	1–15
	How many times did you try to cut down and found you couldn't do so in the past 12 months?	26	3.9 (SD = 2.8)	1–12

^aNumber of respondents asked each question vary depending on their answers to previous questions.

Note: SD, standard deviation.

longer (B), 5% ($n = 5$) of the sample had drunk more frequently than intended and 3% ($n = 3$) had drunk for more days in a row than intended. For *larger/longer (C)*, 1% ($n = 1$) of the sample indicated they had gotten drunk after they had started drinking despite telling themselves they would not. When taking into account the responses for *larger/longer (A)*, *(B)* and *(C)*, 85% of the sample endorsed

the *larger/longer* criterion in the cognitive interview. This is compared with the 93% of the sample who endorsed this criterion in the WMH-CIDI.

Definitional issues: When asked to define “times” for each of the questions, the majority of the respondents indicated that this term referred to single occasions or instances.

Very few of the participants indicated that this term referred to periods of time. For *larger/longer* (A), 77% ($n=77$) defined “times” as occasions/instances. For *larger/longer* (B), 13 of the 21 people asked this question defined “times” as single occasions/instances. For *larger/longer* (C), five of the eight people who were asked this question identified “times” as meaning occasions/instances.

Thresholds for endorsement: Table 3 presents the specific probes as well as the mean number of instances the respondents reported the behaviours related to the *larger/longer* questions in the WMH-CIDI. Given that only one person endorsed *larger/longer* (C), threshold analyses were not possible.

Quit/cut down

Problems from the respondents perspective: Ten per cent ($n=10$) indicated problems with the question for *quit/cut down*. Nine of these 10 people indicated problems with the clarity of the question whilst one person indicated problems with the underlying assumptions of the question.

Double-barrelled question: For *quit/cut down*, 21% ($n=21$) found that they could not stop drinking when they had tried to do so and 26% ($n=26$) reported unsuccessful attempts to cut down on their alcohol consumption. For *quit/cut down*, 34% of the sample endorsed this criterion in the cognitive interview compared with 21% who endorsed this criterion in the WMH-CIDI.

Definitional issues: For *quit/cut down* 56% ($n=56$) identified “times” as referring to single occasions/instances. The respondent definitions of other ambiguous terms in the *quit/cut down* item are listed in Table 2.

Thresholds for endorsement: Table 3 presents these specific probes as well as the mean number of instances the respondents reported the behaviours related to the *quit/cut down* symptoms.

Discussion

Through the novel use of survey methodology and cognitive interviewing, the current study aimed to identify problems in the structure and wording of the WMH-CIDI questions designed to operationalize four DSM-IV diagnostic criteria for AUDs. Using the QAS, a standard tool for questionnaire development, experts in the field identified three overarching problems with the WMH-CIDI questions related to the four target AUD criteria. Three of the criteria were operationalized by questions that were double-barrelled in nature (*tolerance*, *larger/longer* and

quit/cut down). Definitional issues were also identified in each of the questions targeting the criteria. Finally, the explicit thresholds for endorsement of the criteria related to impaired control over alcohol use (*larger/longer* and *quit/cut down*) were also identified as lacking.

Overall, the cognitive interviewing results showed that young adults from the general population experienced some problems in their understanding of the questions designed to assess DSM-IV AUD criteria. When asked during the cognitive interview whether they found any of the questions unclear or confusing, a significant minority indicated that they had issues with the questions. However, none of the respondents sought clarification before providing a response to the WMH-CIDI probes in the initial AUD assessment section. Rather than seek clarification, the respondent may transform ambiguous questions into something they can answer directly without disrupting the standard question-answer exchange in the survey environment. Thus, whilst the respondent appears to provide a reliable response, the validity of this response may be questionable (Chung and Martin, 2005).

Questions remain as to the extent to which the problems identified in the current study are specific to the WMH-CIDI questionnaire or indicative of problems inherent to the DSM-IV text. Almost all of the problems in the WMH-CIDI questions could be directly linked to ambiguities in the DSM-IV text, which is liable to misinterpretation in the absence of clinician judgement. The items in structured diagnostic interviews have been designed to recapitulate the wording of the diagnostic criteria as closely as possible using language that is clear, unambiguous and easily comprehended by the respondent. However, the criterion sets themselves often contain vague terminology that is difficult to operationalize accurately and consistently across different diagnostic measures. The DSM-IV criteria are increasingly being applied to research settings, and it is critical that diagnostic criteria are stated with sufficient objectivity to allow their accurate and consistent operationalization in structured diagnostic interviews (Crowley, 2006). As such, the findings relating to each of the questions, as well as related recommendations for future editions of the DSM, will be discussed in more detail later. It is assumed that clarifications to the DSM-IV text will result in more accurate operationalizations of these criteria in structured diagnostic interviews, thereby leading to more reliable and valid epidemiological data.

Hazard

When the respondents were initially asked whether the question was unclear or confusing, the question targeting

the *hazard* criterion was the most problematic of the WMH-CIDI questions investigated. Post hoc examinations of these responses revealed that several respondents had issues with the examples provided. To increase the applicability of this criterion to younger age groups, alternative examples of hazardous situations may be included in the DSM-IV text, which currently only refers to driving an automobile or operating heavy machinery. When asked to recount the situations they had thought of when answering this question, several respondents mentioned risky sexual behaviours and dangerous pedestrian activity which may be candidates for inclusion in future editions of the DSM.

Many of the DSM-IV alcohol abuse criteria specify that the requisite behaviours need to have occurred “recurrently” and this is operationalized in the WMH-CIDI *hazard* probe by the use of the term “often”. However, “often” was defined in many different ways, indicating this term was interpreted inconsistently by the participants in this study. These results indicate that the term “recurrent” needs to be quantified clearly in DSM-IV to allow for a more precise operationalization of this specifier in diagnostic interviews. Alternative definitions should be proposed and then evaluated empirically, perhaps using item response theory methods to establish maximum discrimination and the most appropriate level of severity in the context of the other criteria.

The level of “substance use” in hazardous situations needs to be operationally defined in the DSM. Does this criterion refer to any substance use, high levels of substance use, or an intermediate level of substance use? Again alternative definitions can be proposed and investigated empirically in the manner described earlier. The WMH-CIDI operationalizes “substance use” as “under the influence” and the majority of respondents in the current study interpreted this as a relatively high level of consumption (i.e. “drunk”). It is currently not clear whether this interpretation is consistent with the intent of the criterion as formulated in the DSM.

Tolerance

The WMH-CIDI question that operationalized *tolerance* was identified as double-barrelled in nature, reflecting the double-barrelled nature of this criterion in the DSM-IV text. This question was less likely to be endorsed when the two questions were decomposed into their respective parts, indicating that the double-barrelled nature of this question may result in some inflation of the prevalence of alcohol dependence amongst young adults in the general population. Future editions of the DSM should

decompose *tolerance* into its respective parts, to enable clearer operationalizations of this criterion in structured diagnostic interviews.

There were also some definitional issues with *tolerance* criterion that need to be resolved. It needs to be clear that the criterion is referring to an increase in the number of drinks consumed, rather than an increase in the strength of the alcoholic beverage consumed. Whilst it may be assumed that an increase in the strength of alcoholic beverage may also indicate an increase in the number of standard drinks consumed, this is not necessarily the case. The terms “buzz” and “high” appear inappropriate in the context of the criterion with respect to alcohol. A significant minority of respondents indicated that they were unable to define these terms in reference to the effects of alcohol, indicating they should be replaced with more contextually relevant terms. “Buzz” and “high” appear to be WMH-CIDI specific operationalizations of the vague term “effect” in the DSM-IV criterion set. A clear definition of the term “effect” (i.e. “intoxicated”) in future editions of the DSM may allow for a more appropriate operationalization of this criterion in diagnostic interviews.

Larger/longer

In the companion paper reporting the results from this study, it was found that the intent of the *larger/longer* criterion was not communicated effectively in the WMH-CIDI largely due to ambiguities in the DSM-IV text. The current results indicate that there are further problems with the questions targeting this criterion in terms of their structure and wording. When the questions for *larger/longer* were broken down into their constituent parts during the cognitive interview, the respondents were less likely to endorse this criterion. As with the *tolerance* criterion, the DSM-IV text describing *larger/longer* is concerned with two independent behaviours that should be listed separately in future editions of the DSM.

In the DSM-IV criterion set, it is stated that the behaviours relating to *larger/longer* need to occur “often” in order for this criterion to be met. However, this term is somewhat ambiguous and has not been operationalized adequately in the diagnostic measures intended to assess AUDs. The current results indicated that, on average, the participants had imposed limits on their drinking several times in the previous year and that they wanted to reduce their consumption by a significant amount. However, the participants also tended to break these limits at quite high rates and when they did not adhere to these self-imposed restrictions, they tended to drink significantly more than intended. The ranges for these variables were also quite

wide, indicating that the number and nature of these restrictions varied considerably, as did the degree with which these limits were broken. This large variability suggests that the *larger/longer* criterion may vary widely with regards to severity. For example, one respondent indicated that they broke their limits once and subsequently endorsed this criterion. Meanwhile, another individual who had endorsed the *larger/longer* criterion reported that they had broken their limit 104 times. Whilst this example demonstrates the wide variability in the behaviours of the respondents who endorsed this criterion, such high rates of these behaviours (as indicated by the latter respondent) also raise questions about the seriousness of the intention to reduce consumption. It appears that the seriousness of the attempt to restrict consumption needs to be adequately communicated in the DSM-IV criterion set, as well as more precise definitions of the term “often”. This term needs to be operationally defined in future editions of the DSM and evaluated in the empirical manner identified earlier (i.e. item response theory to determine appropriate discrimination and severity in the context of the other criteria). The present results indicate that the behaviours relating to this criterion do occur at quite high rates amongst young adults in the general population, providing some guide as to which definitions may be most appropriate.

Quit/cut down

In contrast to *tolerance* and *larger/longer*, more people endorsed the *quit/cut down* criterion when it was broken down into its constituent parts, indicating that the double-barrelled nature of this question may lead to under-endorsement of this criterion amongst young adults in the general population. In order to promote the accurate operationalization of diagnostic criteria, it is important that each criterion is only concerned with a single aspect of behaviour in future editions of the DSM-IV text.

Whilst young adults appeared to understand that this criterion is referring to a long-term and serious intention to cease or reduce alcohol consumption, a significant minority indicated that they thought this question was referring to unsuccessful attempts to control alcohol use on a single occasion. The DSM-IV criterion set needs to be more specific about the nature of the restrictions that are the subject of the *quit/cut down* criterion. It needs to be clearly stated that these restrictions are long-term in nature and not referring to a restriction in consumption relevant to only a single occasion.

Finally, formal thresholds for endorsement of this criterion need to be proposed and tested. As with the

larger/longer criterion, the range with which these behaviours were reported was wide, indicating that this symptom is likely to vary considerably in terms of severity. Future editions of the DSM should be more explicit about the threshold above which this symptom should be reported.

These findings need to be considered in the context of some limitations. Given that only those aged 18–24 were interviewed, it is not possible to determine whether the problems identified in this study are specific to young adults, or also applicable to adolescent and older adult populations. Future research could focus on this important question by replicating the current study amongst a sample with a larger age range. The cognitive interviewing sample was also largely drawn from a university and technical college student population. However, the sample interviewed reported a relatively broad spectrum of drinking frequencies and the rank ordering of AUD criterion endorsement were largely in keeping with that seen in general population surveys (Mewton *et al.*, 2011b, 2011c). Due to time constraints imposed by the intensive nature of the cognitive interviewing technique, it was only feasible to investigate problems associated with four out of the 11 AUD criteria. These criteria were chosen based on previous empirical investigations of potential criterion bias. However, it remains to be investigated whether some of the other criteria, particularly the new *craving* criterion proposed for DSM-5 (American Psychiatric Association, 2010), are performing as intended. Finally, the cognitive interviews were conducted retrospectively, after the WMH-CIDI questionnaire had been administered. It is possible that concurrent administration of the cognitive interview and the WMH-CIDI may have produced different results and prompted a more naturalistic exchange between the interviewer and respondent. However, independent administration of the WMH-CIDI allowed assessment of alcohol abuse and dependence in this sample, which was necessary for the primary aims of the project as reported previously (Slade *et al.*, 2012).

In conclusion, the present study used novel investigative techniques informed by survey methodology to investigate ambiguous terminology in four of the DSM-IV AUD criteria. Systematic expert reviews identified problems in the structure and wording of the DSM-IV AUD criteria as operationalized by the WMH-CIDI. Many of these problems were the result of ambiguities in the DSM-IV AUD text which need to be addressed to ensure consistent operationalizations of the criteria. This will allow more accurate prevalence estimates in surveys based lay administered structured diagnostic interviews. The results from this study were translated into empirically-

based recommendations for modifications to the DSM-IV AUD criteria, demonstrating the ways in which survey methodology can be implemented as a means of improving and revising diagnostic criteria. Whilst some changes have been proposed for DSM-5 AUD, the structure and wording of the symptoms remain consistent between DSM-IV and DSM-5 (American Psychiatric Association, 2010). The current study indicates problems in this structure and wording and suggests ways in which the

symptom text may be changed to enable more accurate operationalizations of the criteria in structured diagnostic interviews.

Acknowledgement

This manuscript was supported by an Australian National Health and Medical Research Council Project Grant (#630414).

References

- Agrawal A., Heath A.C., Lynskey M.T. (2011) DSM-IV to DSM-5: the impact of proposed revisions on diagnosis of alcohol use disorders. *Addiction*, **106**(11), 1935–1943.
- American Psychiatric Association (2010) DSM-5 Development. <http://www.dsm5.org/Pages/Default.aspx> [24 May 2010]
- Beatty P.C., Willis G. (2007) Research synthesis: the practice of cognitive interviewing. *Public Opinion Quarterly*, **71**(2), 287–311.
- Caetano R. (1999) The identification of alcohol dependence criteria in the general population. *Addiction*, **94**(2), 255–267.
- Chung T., Martin C.S. (2002) Concurrent and discriminant validity of DSM-IV symptoms of impaired control over alcohol consumption in adolescents. *Alcoholism: Clinical & Experimental Research*, **26**(4), 485–492.
- Chung T., Martin C.S. (2005) What were they thinking? Adolescents' interpretations of DSM-IV alcohol dependence symptom queries and implications for diagnostic validity. *Drug & Alcohol Dependence*, **80**(2), 191–200.
- Chung T., Martin C.S., Winters K.C., Langenbucher J.W. (2001) Assessment of alcohol tolerance in adolescents. *Journal of Studies on Alcohol*, **62**(5), 687–695.
- Chung T., Martin C.S., Armstrong T.D., Labouvie E.W. (2002) Prevalence of DSM-IV alcohol diagnoses and symptoms in adolescent community and clinical samples. *Journal of the American Academy of Child & Adolescent Psychiatry*, **41**(5), 546–554.
- Crowley T.J. (2006) Adolescents and substance-related disorders: research agenda to guide decisions on Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-V). *Addiction*, **101**(Suppl. 1), 115–124.
- Edwards G., Gross M.M. (1976) Alcohol dependence: provisional description of a clinical syndrome. *British Medical Journal*, **1**(6017), 1058–1061.
- Harford T., Yi H., Faden V., Chen C. (2009) The dimensionality of DSM IV alcohol use disorders among adolescent and adult drinkers and symptom patterns by age, gender, and race/ethnicity. *Alcoholism: Clinical and Experimental Research*, **33**(5), 868–878.
- Kelley K., Clark B., Brown V., Sitzia J. (2003) Good practice in the conduct and reporting of survey research. *International Journal for Quality in Health Care*, **15**(3), 261.
- Kessler R.C., Ustun T.B. (2004) The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *International Journal of Methods in Psychiatric Research*, **13**(2), 93–121.
- Martin C.S., Chung T., Kirisci L., Langenbucher J.W. (2006) Item response theory analysis of diagnostic criteria for alcohol and cannabis use disorders in adolescents: implications for DSM-V. *Journal of Abnormal Psychology*, **115**(4), 807–814.
- Mewton L., Slade T., McBride O., Grove R., Teesson M. (2011a) An evaluation of the proposed DSM 5 alcohol use disorder criteria using Australian national data. *Addiction*, **106**(5), 941–950.
- Mewton L., Teesson M., Slade T., Cottler L. (2011b) The psychometric performance of DSM-IV alcohol use disorders in young adulthood: evidence from an Australian general population sample. *Journal of Studies on Alcohol and Drugs*, **72**, 811–822.
- Mewton L., Teesson M., Slade T., Grove R. (2011c) The epidemiology of DSM-IV alcohol use disorders amongst young adults in the Australian population. *Alcohol and Alcoholism*, **46**(2), 185–191.
- Slade T., Teesson M., Mewton L., Memedovic S., Krueger R. F. (2012) Do young adults interpret the DSM diagnostic criteria for alcohol use disorders as intended? A cognitive interviewing study. *Alcoholism: Clinical and Experimental Research*, **37**(6), 1001–1007.
- Vergés A., Littlefield A.K., Sher K.J. (2011) Did lifetime rates of alcohol use disorders increase by 67% in 10 years? A comparison of NLAES and NESARC. *Journal of Abnormal Psychology*, **120**(4), 868–877.
- Willis G. (2005) Cognitive Interviewing: A Tool for improving Questionnaire Design, Thousand Oaks: CA, Sage Publications.
- Willis G., Lessler J. (1999) Question Appraisal System Manual, Raleigh, NC: Research Triangle Institute.