Interpretation of DSM-5 alcohol use disorder criteria in self-report surveys may change with age. A longitudinal analysis of young Swiss men

Simon Marmet[a], Joseph Studer[a], Nicolas Bertholet[a], Véronique S. Grazioli[a], Jean-Bernard Daeppen[a] & Gerhard Gmel[a,b,c,d]

a Alcohol Treatment Centre, Lausanne University Hospital/CHUV, Switzerland (Avenue de Beaumont 21b, 1011 Lausanne)

b Addiction Switzerland, Lausanne, Switzerland (Av. Louis-Ruchonnet 14, 1001 Lausanne)

c Centre for Addiction and Mental Health, Toronto, Canada (1001 Queen Street West, Toronto ON - M6J 1H4)

d University of the West of England, Bristol, United Kingdom (Frenchay Campus, Coldharbour Lane, Bristol BS16 1QY)

Corresponding Author:

Simon Marmet

Service d'Alcoologie

P2.01.113

Avenue de Beaumont 21b

CH-1011 Lausanne

Phone: +41 (0) 21 314 18 97

E-mail: simon.marmet@chuv.ch

Abstract

Background

General population surveys using self-reported measures show alcohol use disorder (AUD) to be most prevalent in young adulthood, but this may be due to misinterpretations of AUD criteria among inexperienced drinkers. The present study investigated changes in prevalence of criteria during emerging adulthood.

Methods

4924 young Swiss men were followed across three waves (age at wave 1 (w1): 20; w2: 21; w3: 25 years). We measured AUD according to the 11 DSM-5 criteria and estimated Item Response Theory models for each wave and differential item functioning across waves, related to the cohort growing older.

Results

Endorsement of several AUD criteria varied considerably as a function of age in a period of only five years: Five criteria showed differential item functioning between waves 1 and 3 (i.e. between the age of 20 and 25), including the three most frequently endorsed criteria. Prevalence of *tolerance* (w1, 57.8%; w3, 29.6%) decreased in relation to the AUD score (Mantel–Haenszel OR=0.26), whereas the use of alcohol *larger/longer* than intended (w1, 73.7%; w3, 84.8%; OR=1.93) and *hazardous use* (w1, 62.7%; w3, 68.4%; OR=1.31) increased, resulting in an unchanged mean AUD score and prevalence, but changing combinations of criteria to qualify AUD.

Conclusions

Considerable differential item functioning over five years among several of the most endorsed AUD criteria suggests shifts in the interpretation of DSM-5 criteria during emerging

3

adulthood. Self-reported measures of DSM-5 AUD criteria may require reformulation to account for young people's different perceptions and to yield comparable diagnoses over time and across age groups.

1. Introduction

Alcohol use causes much harm: it is among the most prevalent causes of death among young men in Switzerland (Marmet et al. 2014). A general population survey based on self-reported measures showed the highest prevalence of alcohol use disorder (AUD) to be among young Swiss men aged 15–24 years old (Marmet and Gmel 2014). Similar tendencies for alcohol dependence have been found in general population surveys in Germany (Pabst et al. 2012), the USA (Hasin et al. 2007) and Australia (Teesson et al. 2010). This is in stark contrast to the traditional idea that AUD, and particularly the symptoms of physiological dependence (withdrawal, tolerance and craving), develops over decades of alcohol consumption (Langenbucher and Chung 1995; Caetano and Babor 2006; Winters et al. 2011). In this view, AUD should be rather rare in young adults, who have a relatively short history of drinking. Furthermore, Switzerland's AUD treatment and hospitalisation statistics peak around the age of 50 (Astudillo and Maffli 2017; Marmet et al. 2017), with few young adults being treated for AUD. Some researchers assume that the very high prevalence of AUD in general population surveys is due to false-positives (Regier et al. 1998; Narrow et al. 2002; Wakefield and Schmitz 2014), particularly among younger age groups. A recent review by Lane et al. (2016) also showed that endorsements of AUD criteria varied according to sample type (e.g. representative surveys vs clinical assessment) and age. For epidemiological research, it is important to know whether perceptions of the properties of items/criteria change with age and what impact this has on prevalence estimates for AUD (Harford et al. 2005). Using the framework of Item Response Theory (IRT), the present study analyses changes in perceptions of AUD criteria, with respect to the scale score, from early to emerging adulthood in a longitudinal study.

Several studies to date have provided evidence that some of the DSM-5 (American Psychiatric Association 2013) AUD criteria might be misunderstood and that their perception or interpretation may even change with age, i.e. young people may answer questions assessing AUD criteria differently from older people and not as was intended by the DSM (Harford et al. 2005; Caetano and Babor 2006; Winters et al. 2011; Pabst et al. 2012). For example, physiological *tolerance* (i.e. quantity of alcohol which must be consumed for the same effect; see sample question-wording in Table 1) is supposed to increase with age in younger years because of repeated exposure to alcohol and increasing body mass; however, in young adulthood, *tolerance* is actually reported most often by inexperienced drinkers, and subsequently less often after the age of about 23 (Chung et al. 2001; Harford et al. 2005).

Table 1 about here

Similarly, *withdrawal* (i.e. negative effects after abstinence from alcohol) has been found to be most often reported among young people, which may be due to misunderstanding the sequelae of an acute intoxication (i.e. a "hangover" on the day after alcohol consumption) as a symptom of withdrawal (Caetano and Babor 2006). Karriker-Jaffe et al. (2015) monitored young to middle-aged individuals and found that 28% of those reporting withdrawal were actually reporting the aftereffects of heavy episodic drinking and not physiological *withdrawal*. Misunderstandings may also be related to the different wordings used in survey questions. In some measurement instruments, *withdrawal* must have occurred at least three times or during an extended period, in others, at least once with no stated duration (Lane et al. 2016), which could have a significant impact on symptom prevalence.

6

The *larger/longer* criterion (i.e. drank larger quantities of alcohol than intended or longer than intended) may be more strongly related to social pressure, particularly among young people, than to a compulsion to drink (Wakefield and Schmitz 2014). For example, Karriker-Jaffe et al. (2015) showed that about 50% of individuals endorsing the *larger/longer* criterion did so because of social pressures to drink rather than a compulsion to drink. The *larger/longer* criterion was also found to be more often reported by younger age groups than by older ones (Harford et al. 2005; Pabst et al. 2012).

Hazardous use (e.g. drink-driving), on the other hand, a DSM-IV alcohol abuse criterion, may be more frequent among young adults than adolescents because of the broader availability of motorised vehicles. Greater availability increases the likelihood of finding oneself in a dangerous situation while driving under the influence of alcohol (Winters et al. 2011). Additionally, due to its significant dependence on drink-driving, this criterion is associated with socioeconomic status (i.e. access to automobiles) (Babor and Caetano 2008), but also to environmental (stronger enforcement of drink-driving laws) and cultural issues (Martin et al. 2011).

Given these variations in the interpretation of criteria across age groups, it may be difficult to distinguish changes in the severity of AUD over time due to the problems related to alcohol use and the *typical* developmental processes associated with the transition from adolescence to young adulthood. Accordingly, Harford et al. (2005) concluded that the higher prevalence of some DSM-IV alcohol dependence criteria among adolescents and young adults might blur the distinction between DSM-IV alcohol dependence and normative development of drinking patterns (see also Slade et al. (2013)).

A closer look at the AUD criteria is also of interest in light of the change from DSM-IV to DSM-5. DSM-5 collapsed the dependence (7 criteria) and abuse categories (4 criteria) from

DSM-IV into AUD measured with 11 criteria, removing the *legal* criterion and adding the *craving* criterion. Furthermore, DSM-5 now differentiates between mild, moderate and severe AUD, with the cut-off for mild AUD set at 2 criteria. This means that 2036 different combinations of criteria can equate to *at least mild* AUD (Gmel 2015). Some of these combinations may consist of two relatively mild criteria, which may not adequately reflect the definition of compulsive substance use (Winters et al. 2011).

1.2   Aims

Using an Item Response Theory (IRT) framework, this study aimed to assess which AUD criteria were most often endorsed by young people and to test changes in criterion prevalence across time using differential item functioning. IRT studies enable an assessment of the properties of individual items (respectively criteria) with respect to a latent trait (i.e. severity of AUD) and how this changes across time. To date, the few IRT studies on AUD criteria have mostly been conducted in the USA, thus limiting any possible generalisation of meta-analytic findings (Lane et al. 2016). We also examined whether changes in the responses to DSM-5 criteria were indeed related to participants' ages or rather to the time since onset of AUD.

2. Methods

2.1 Sample

The present work's sample population came from the Cohort Study on Substance Use Risk Factors (C-SURF; see Gmel et al. (2015) for an overview), a longitudinal study designed to examine use patterns and associated factors among young Swiss men. Enrolment for the baseline measurement occurred between August 2010 and November 2011 in three of Switzerland's six military recruitment centres, located in Lausanne (French-speaking: 57.4% of the final sample), Windisch and Mels (German-speaking: 42.6%), during the military

recruitment procedures which are mandatory for all Swiss men. A total of 13'237 young men were informed and invited to enrol in the study. Among them, 7556 (57.1%) gave written consent to be contacted for participation in the study (for more details, see Gmel et al. (2015); Studer et al. (2013)). Overall, 5987 (79.2%) young men completed the baseline questionnaire (wave 1, w1) between September 2010 and March 2012; of those, 4924 (82.2%) completed the two follow-up questionnaires at an average of 15.8 months (wave 2, w2; March 2012–January 2014) and 65.2 months (wave 3, w3; March 2016–November 2017) after the baseline questionnaire. The final sample of 4924 included all the participants who had replied to all three waves (i.e. 82.2% of those who replied to the baseline questionnaire). The research protocol was approved by the Human Research Ethics Committee of the Canton Vaud (Protocol No. 15/07).

2.2 Measurements

AUD (last 12 months) was measured using DSM-5's (American Psychiatric Association 2013) 11 AUD criteria. Items from Knight et al. (2002) and additional *craving* criteria from Grant et al. (2003) were translated into French and German. Participants were asked to indicate whether they had experienced each criterion in the previous 12 months (see Table 1). The *hazardous use* criterion was measured using two questions, of which at least one had to be affirmed for the criterion to be fulfilled. Missing values on items (ranging between 9 and 17 cases per item) were replaced by the scale mean. individuals with missing values on more than 20% of their items were excluded from analysis using the dichotomous AUD diagnosis (12 cases in w1, 12 in w2 and 9 in w3), but were included in criterion-wise analysis.

Alcohol use was measured in drinks-per-day using an extended quantity–frequency questionnaire measuring frequency and quantity separately on workweek days and weekend days. This measure was shown to be more strongly associated with consequences of alcohol

use and AUD than a simple quantity–frequency measure or past-week's-drinking diary (Gmel et al. 2014). Heavy episodic drinking was measured using one question about how often participants consumed six or more standard drinks on one occasion, and the maximum number of standard drinks in the past 12 months was determined using an open-ended question. Images of standard drinks (e.g. a glass of beer or wine), corresponding to about 10 grams of pure alcohol, were included in the questionnaire for reference.

2.3 Statistical Analysis

Differences in the prevalence of AUD between waves were tested for significance using McNemar's chi-squared tests in SPSS 25 software; differences between AUD scale means were tested using paired sample t-tests. IRT models for all three waves were calculated in Stata 15 software using two-parameter models (difficulty and discrimination) for binary data (Lord and Novick 2008). Briefly, difficulty is defined as the threshold on a latent variable for the severity of AUD (theta) where the probability of endorsing a criterion is 50%, i.e. an *easy* item is endorsed by a larger proportion of the sample than a *difficult* one. Discrimination describes the criterion's ability to discriminate between individuals with different symptom severity (Baker 2001). Differential item functioning (DIF) across waves was assessed using Stata 15's Mantel–Haenszel procedure. This tests whether an item's difficulty changes between two groups (in this case, between two waves) with respect to the latent scale score.

Our approach measuring differential item functioning across waves, i.e. across age, implies that the presence of individual criteria varies as a function of age of the participants. However, changes in the prevalence of criteria may also be influenced by the time of onset of AUD, i.e late onsetters being older and having other combinations than early onsetters. To determine whether the prevalence of criteria depended more on the participant's age or on the time since onset of AUD, participants with AUD in w2 or w3 were classified either as *incident* cases,

with a first occurrence of AUD in the relevant wave, or as *persistent* cases, if the AUD of the corresponding wave had already occurred before in other waves. For example, someone with no AUD in wave 1, but AUD in wave 2, was classified as *incident* in wave 2. To the contrary, someone with AUD in wave 1 and wave 2 would have been classified as having a *persistent* AUD in wave 2. Participants with AUD in w1 and w3, but who did not meet the threshold for AUD in w2, where classified as *persistent* in wave 3, because they showed already AUD before and – despite remission in w2 – had a recurrent AUD in wave 3. Individuals with AUD in all three waves were also classified as *persistent* in wave 2 and wave 3, but incident in wave 1. If changes in the prevalence of criteria across time were mostly due to age, few differences should be expected between *incident* and *persistent* cases, given that both groups were of about the same age. If changes in the prevalence of criteria were due to time since the onset of an AUD, differences in criterion endorsement between *incident* and *persistent* cases should be expected. Logistic regressions in SPSS 25 were used to determine whether there were differences in criterion prevalence between incident and persistent cases in w2 and w3. These regressions were also adjusted for the total AUD score to account for differences in the severity of AUD between the two groups. Because all cases with AUD in wave 1 were classified as incident in wave 1, this anaylsis was not possible for wave 1. Independent-samples *t*-tests showed no significant differences in age between the incident and persistent cases, therefore the regressions were not adjusted for age.

3.   Results

A total of 31.1% of participants reported at least a mild AUD at w1 (mean age 19.97 years), 31.8% at w2 (mean age 21.30 years) and 31.9% at w3 (mean age 25.40 years; see Table 2); thus overall prevalence did not change across the three waves (p = .309 between w1 and w3), and therefore with age (wave being a proxy for age). Between w1 and w3, i.e. between the age of about 20 to about 25, there was a significant increase in mild AUD (21.3% to 23.2%; p = .011) and a significant decrease in severe AUD (3.1% to 2.1%; p = .010). Alcohol use quantity, i.e. heavy volume, (3 drinks or more per day, p < .001; maximum number of drinks, p < .001 and prevalence of monthly heavy episodic drinking, p < .001) decreased from w1 to w3. Of those with at least mild AUD at w1, 59.2% also had AUD at w2, and 40.0% had AUD at w2 and w3. A total of 53.2% of participants reported at least mild AUD in at least one of the three waves. Cronbach's alphas for AUD in the three waves were .738 at w1, .711 at w2 and .714 at w3, indicating acceptable internal consistency (Gliem and Gliem 2003).

Table 2 about here

Table 3 presents results from the IRT analysis of the total sample (presented graphically for baseline in Figures 1 and 2). Item difficulties (respectively prevalence rates) varied widely at baseline, from *larger/longer* at a difficulty of 0.74 (prevalence 29.6%) to *quit/control* at 2.29 (prevalence 3.4%). Broadly, two groups of items were distinguished based on difficulty: five *easy* ones (the five separated lines on the left of Figure 1) with difficulties below 2 (prevalence above 10%), namely *larger/longer*, *tolerance*, *hazardous use*, *time spent* and *neglect roles*. The six remaining items were relatively more difficult, with difficulties between 2.17 and 2.29 at baseline and prevalence between 2.6% and 4.0%, respectively. According to the classification developed by Baker (2001), discrimination was moderate for *hazardous use* (1.26 at w1, and an even lower 1.07 at w3), and high or very high (> 1.70) for the other items in all three waves.

Table 3 about here

Figures 1 and 2 about here

Table 4 reports the prevalence of criteria and the cumulative percentage of items among participants with at least mild AUD. Among participants with at least mild AUD in w1, 73.7% endorsed the *larger/longer* criterion and 62.7% the *hazardous use* criterion. If only these two criteria had been measured instead of the full 11-criteria scale, 42.4% (56.5% in w3) of the participants with any type of AUD would nonetheless have been diagnosed with mild AUD (at least two criteria). Similarly, 94.8% of participants with at least mild AUD in w1 would have received the same diagnosis if only the first five items had been used. Thus, the last six items (Table 3) were not highly relevant to the diagnosis of at least mild AUD. In fact, only 203 cases (13.3% of these with a mild AUD diagnosis) at wave 1 showed at least two of the six last difficult items (result not shown in tables).

Table 4 about here

Five criteria showed differential item functioning between w1 and w3 (Table 3). The difficulty of *tolerance* increased between w1 and w3, but this was to a large part compensated for by lower difficulties in w3 in the frequently endorsed items of *larger/longer* and *hazardous use*, and to a minor extent *quit/control* and *craving*, resulting in an overall unchanged mean score for the AUD scale between w1 and w3 (p = .892; see Table 1).

To investigate whether changes in the prevalence of items were due to increasing age or the time since the onset of AUD, we tested the differences between incident and persistent cases of AUD in w2 and w3 (see Table 5). There were no significant differences between the incident and persistent cases in w3 (and only a few between w1 and w2) when adjusted for differences in the severity of AUD between the two groups. This indicated that the prevalence of the criteria depended mostly on the age of the participant and less on the time since the onset of an AUD.

Table 5 about here

4.   Discussion

4.1 Prevalence of AUD

Using self-reported DSM-5 criteria, almost one third (w1 (mean age 19.97 years): 31.1%; w2 (mean age 21.30 years): 31.8%; w3 (mean age 25.40 years): 31.9%) of participants endorsed criteria for at least mild AUD in each of the three waves. These prevalence rates are high compared to those from a general population survey in Switzerland (5.7% of the general population and 7.8% of men), but the same survey also reported that the highest prevalence of DSM-5 AUD was among younger age groups (Marmet and Gmel 2014). This is consistent with surveys in Germany (Pabst et al. 2012), the USA (Hasin et al. 2007) and Australia (Teesson et al. 2010), which reported high rates of DSM-IV alcohol dependence in younger age groups.

4.2 IRT properties of items and criterion combinations

As regards individual criteria, they could be broadly separated into two groups: five items were rather easy in terms of IRT (*larger/longer*, *tolerance*, *hazardous use*, *time spent*, *neglect roles*), and six were rather difficult (*social/interpersonal problems*, *withdrawal*, *quit/control*, *craving*, *activities given up*, *physical/psychological problems*). Regarding combinations of criteria, it is noteworthy that more than 90% of participants with at least mild AUD would still have had this diagnosis if only the first five *easy* items had been used. On the other hand, only 13.3% of participants with mild AUD in wave 1 would have received this diagnosis if only the last six *difficult* items had been used. It is debatable whether combinations of such *easy* criteria, in the absence of important markers of AUD like withdrawal or use despite physical/psychological problems, are a good representation of the construct of AUD. Given

the predominance of combinations of easy items, one could also ask whether more than 30% prevalence of AUD, even if labelled "mild", is a reasonable estimate for AUD in young men.

4.3 Longitudinal changes in item difficulties

The present study shows that the difficulty of some criteria (and thus the prevalence of endorsement) changes rapidly between 20 and 25 years old. IRT analysis showed that five items showed significant differential item functioning with age, i.e. between w1 (mean age 19.97 years) and w3 (mean age 25.40 years). The largest change in difficulty (from 1.11 in w1 to 1.80 in w3) was found for *tolerance*, with a DIF measured using the Mantel–Haenszel OR of 0.26. Among participants with at least mild AUD, *tolerance* was prevalent in 57.8% in w1, but only 29.6% in w3. This decrease in the prevalence of *tolerance* was mostly counterbalanced by an increase in the prevalence of the *larger/longer* and *hazardous use* criteria, resulting in an unchanged overall mean score and prevalence for AUD. Thus, the functioning of several items changed fundamentally with respect to the scale score within only 5 years. Importantly, these changes affected the three most prevalent criteria which were highly relevant for a diagnosis of mild AUD. Indeed, almost three quarters of the mild AUD diagnoses would nonetheless have been made if only these three items had been used instead of the full scale (Table 4). In our sample, the prevalence of at least mild AUD was stable across waves (between 31% and 32%), yet only 40% of those with AUD at w1 endorsed AUD in all three waves, meaning that there were a lot of changes in AUD status, including remissions and an equal number of new onsets. This frequent changes are also visible in the fact that more than half (53.2%) of our sample reported at least mild AUD in at least one of the three waves. In the light of our findings, it is difficult to determine whether these individual changes in AUD status reflected a true change in AUD severity (i.e. remission or onset) or whether they are due to a different understanding of AUD criteria at 25 years old than at 20 (i.e. measurement error).

15

One major advantage of the present study is its longitudinal design using unchanged instruments. The changes in prevalence found can largely be attributed to age effects and not to the effects of time since onset of AUD. This was shown by testing differences in the prevalence criteria for participants with incident AUD versus participants with persistent AUD, which were not significant for most criteria.

4.4 Conceptual and empirical findings concerning critical age-dependent criteria

The three criteria (*tolerance*, *larger/longer* and *hazardous use*) showing the most significant changes in our study are among the most disputed in the literature as regards bias in samples of young peoples (Harford et al. 2005; Caetano and Babor 2006; Winters et al. 2011; Pabst et al. 2012). One of the problems with tolerance is the unspecified period in which the change in alcohol use took place, as well as a lack of precision regarding the amount of change in alcohol use (e.g. 50% higher in earlier versions of the DSM). Young people are often in the early stages of their drinking lives and thus are experimenting with increasing amounts of alcohol at this time. However, this is more related to the developmental trajectories of increased drinking experiences than to physiological tolerance (Slade et al., 2013), something which commonly develops after a long period of chronic exposure to alcohol. Thus, men aged 20 have a better recollection of increases in the quantity of alcohol consumption compared to when they began drinking, and they therefore frequently report greater tolerance. Aged 25, however, alcohol consumption may have stabilised or may even have decreased, therefore resulting in a lower perceived tolerance (Winters et al. 2011). Harford et al. (2005) showed similar effects with an increasing prevalence of *tolerance* from adolescence to early adulthood (peaking at around 22), followed by decreases in prevalence after the age of 23. A German study also reported that *tolerance* was more frequent in younger people (aged 18–24) than in older age groups (Pabst et al. 2012). Thus, our finding of a lower prevalence of reported *tolerance* with age is well in line with precedent findings.

One explanation for the increase in the prevalence of the *larger/longer* criterion could be that some sort of drinking limit has been set, something rarely done in adolescence (Chung and Martin 2005) but more frequently at the age of 25 because certain role obligations (e.g. family, job) require restrictions in drinking behaviour. However, contrary to our findings, Harford et al. (2005) found that the prevalence of the *larger/longer* criterion in men decreased after the age of 23, and Pabst et al. (2012) found that *larger/longer* was reported more frequently in the 18–24-year old age group than in older age groups. However, our study covers a relatively short time span and it is possible that the prevalence of *larger/longer* will decrease in the future. The *larger/longer* criterion has also been criticised because it may mostly reflect peer pressure in social environments, which would not be a sign of alcohol use disorder per se. Karriker-Jaffe et al. (2015) estimated that excluding individuals reporting *larger/longer* drinking for social reasons would decrease AUD prevalence estimates by 8%. Similarly, Slade et al. (2013), using cognitive interviewing, found that a large majority (70.6%) reported continuing drinking for social and non-compulsive reasons.

The increase in the prevalence of *hazardous use* may be explained by more participants travelling by car in w3, and therefore being more often in danger when returning home after having drunk alcohol (Winters et al. 2011). The validity and utility of this criterion is another hotly debated issue (Babor and Caetano 2008; Martin et al. 2011), particularly for the US, where it is mostly related to drink-driving (Babor and Caetano 2008). As shown by other authors, *hazardous use* is heavily influenced by social circumstances (Martin et al. 2011) and perhaps also by cultural interpretations. For example, in Switzerland, the maximum blood-alcohol level for newly licenced drivers is 0.05 mg/l (0.01%, corresponding to less than one drink; Swiss Federal Chancellery (2018)), which may lead to relatively small amounts of alcohol being perceived as dangerous.

Differences in cultural interpretations may also play a role with the *withdrawal* criterion. Whereas *withdrawal* is prevalent in US samples and often seems to be misinterpreted as the symptoms of a hangover after a heavy drinking session rather than the physiological effects of AUD (Caetano and Babor 2006; Karriker-Jaffe et al. 2015), it was not highly prevalent in the present study, confirming earlier findings from Germany (Pabst et al. 2012). Besides cultural differences in the interpretation of this criterion (Rehm and Room 2015), differences between studies may be due to subtle differences in question-wording. In some instruments, the symptoms of *withdrawal* must have occurred at least three times or during an extended period (of more than a day, as in the present study), which may cause a misinterpretation of rare events of acute intoxication, whereas other instruments use a definition of at least once without a duration (Lane et al. 2016).

In response to the high prevalence of AUD in general population surveys, possibly due to misunderstandings and over-reporting of some criteria, Wakefield and Schmitz (2015) suggested a harm–dysfunction approach, in which criteria that implicate neither harm nor dysfunction (i.e. *tolerance*, *hazardous use*, *larger/longer*) were to be excluded. Interestingly, this were the three criteria which the present study found to be the most influenced by age effects. Wakefield and Schmitz (2015) suggested further that at least one harm item (e.g. *activities given up*, *neglect roles*) and one dysfunction item (e.g. *withdrawal*, *quit/control*) should be present before diagnosing AUD. According to Wakefield and Schmitz, their approach eliminated almost 83% of their so-called *teen transient false-positive group* in their sample. Their approach would be similar to only assigning a diagnosis of AUD if at least two of the six difficult items in our study were required to be present, which would in the case of our sample at wave 1 eliminate 86.7% of all diagnosis with at least mild AUD.

Our results are limited to a sample of young Swiss men, but they are of particular relevance because this is the group with the highest levels of self-reported AUD in general population

surveys. Changes in the endorsements of the AUD criteria between the ages of about 20 and 25 indicate a considerable age bias in the estimation of AUD in self-reported surveys. Therefore, it cannot be assumed that estimates for AUD in young people can be compared with AUD in other age groups, e.g. people around the age of 50 who are in the most frequently treated and hospitalised group for AUD (Astudillo and Maffli 2017; Marmet et al. 2017) despite lower reported AUD in general population surveys. Comparing AUD prevalence rates from self-reported general population surveys between young people and older age groups may therefore need a high degree of scrutiny. Indeed, AUD in different age groups may reflect different problems, i.e. AUD in young people may represent a type of *youthful* AUD (Nelson and Wittchen 1998; Caetano 1999; Caetano and Babor 2006) which is distinct from the AUD seen in clinical samples. This *youthful* AUD could be described as a rather unstable condition, often disappearing with time and without treatment.

Nevertheless, although most of the AUDs identified in young men via self-reported general population surveys may be qualitatively distinct from the AUDs in older age groups or in clinical samples involving professional diagnoses, they may nevertheless cause individual suffering. Furthermore, young people with AUD may be at a higher risk for developing into more serious AUDs or other conditions later on (Rohde et al. 2001). They may therefore require professional attention, although perhaps in a different form from than given to older people with AUD. The possibility that most of the AUDs in young men measured by self-report questionnaires may be types of *youthful* AUD does also not imply that severe and clinically relevant forms of AUD do not exist among that age group.

4.5 Conclusion

In the age group (aged from 20 to 25) most frequently affected by alcohol use disorder (AUD) as assessed with self-reported questions for the DSM-5 criteria, these criteria showed poor

psychometric properties and poor stability as regards item functioning in the longitudinal perspective. With regards to research and treatment, it should be kept in mind that self-reported measures of AUD may not measure the same thing across an individual's life, and a diagnosis of AUD should be interpreted differently for different age groups. A more precise formulation for the DSM-5 criteria could be valuable, making them less vulnerable to misinterpretation (see (Chung et al. 2001; Karriker-Jaffe et al. 2015). Also, some criteria might need rewording to allow for a better spread of criteria across the difficulty spectrum (Mewton et al. 2013). The DSM-5 cut-off for mild AUD should be used with caution, especially among young men, as many of these mild AUD diagnoses are based on items of questionable validity among adolescents, who may easily misunderstand them (Caetano and Babor 2006; Winters et al. 2011), and which also are interpreted differently across age groups. Furthermore, using the mild AUD cut-off would suggest an epidemic of AUD among young Swiss men, which is, in the opinion of the authors, at variance with reality.

5.   References

American Psychiatric Association. 2013. Diagnostic and statistical manual of mental disorders (5th ed.). Washington, DC: Author: American Psychiatric Publishing.

Astudillo M, Maffli E. 2017. Ambulante Suchthilfe - Ergebnisse der KlientInnenbefragung 2015. Lausanne: Sucht Schweiz.

Babor TF, Caetano R. 2008. The trouble with alcohol abuse: what are we trying to measure, diagnose, count and prevent? Addiction. 103(7):1057-1059.

Baker FB. 2001. The basics of item response theory. ERIC.

Caetano R. 1999. The identification of alcohol dependence criteria in the general population. Addiction. 94(2):255-267.

Caetano R, Babor TF. 2006. Diagnosis of alcohol dependence in epidemiological surveys: an epidemic of youthful alcohol dependence or a case of measurement error? Addiction. 101(s1):111-114.

Chung T, Martin CS. 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 CS, Winters KC, Langenbucher JW. 2001. Assessment of alcohol tolerance in adolescents. Journal of Studies on Alcohol. 62(5):687-695.

Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. 2003: Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education.

Gmel G. 2015. ICD-11: Die Trennung von der Forschung: Applaus der Kliniker. [Dissociation from Research: Applause by clinicians [commentary]. SUCHT. 61(5):123-125.

Gmel G, Akre C, Astudillo M, Bähler C, Baggio S, Bertholet N, Clair C, Cornuz J, Daeppen J-B, Deline S et al. 2015. The Swiss Cohort Study on Substance Use Risk Factors – Findings of two waves. SUCHT. 61(4):251-262.

Gmel G, Studer J, Deline S, Baggio S, N'Goran A, Mohler-Kuo M, Daeppen J-B. 2014. More is not always better—comparison of three instruments measuring volume of drinking in a sample of young men and their association with consequences. Journal of studies on alcohol and drugs. 75(5):880-888.

Grant BF, Dawson DA, Stinson FS, Chou PS, Kay W, Pickering R. 2003. The Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (AUDADIS-IV): reliability of alcohol consumption, tobacco use, family history of depression and psychiatric diagnostic modules in a general population sample. Drug and alcohol dependence. 71(1):7-16.

Harford TC, Grant BF, Yi Hy, Chen CM. 2005. Patterns of DSM-IV alcohol abuse and dependence criteria among adolescents and adults: Results from the 2001 National Household Survey on Drug Abuse. Alcoholism: Clinical and Experimental Research. 29(5):810-828.

Hasin DS, Stinson FS, Ogburn E, Grant BF. 2007. Prevalence, correlates, disability, and comorbidity of DSM-IV alcohol abuse and dependence in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. Archives of general psychiatry. 64(7):830-842.

Karriker-Jaffe KJ, Witbrodt J, Greenfield TK. 2015. Refining measures of alcohol problems for general population surveys. Alcoholism: clinical and experimental research. 39(2):363-370.

Knight JR, Wechsler H, Kuo M, Seibring M, Weitzman ER, Schuckit MA. 2002. Alcohol abuse and dependence among U.S. college students. Journal of Studies on Alcohol. 63:263-270.

Lane S, Steinley D, Sher K. 2016. Meta-analysis of DSM alcohol use disorder criteria severities: structural consistency is only 'skin deep'. Psychological medicine. 46(8):1769-1784.

Langenbucher JW, Chung T. 1995. Onset and staging of DSM-IV alcohol dependence using mean age and survival-hazard methods. Journal of Abnormal Psychology. 104(2):346.

Lord FM, Novick MR. 2008. Statistical theories of mental test scores. IAP.

Marmet S, Gmel G. 2014. Suchtmonitoring Schweiz –Alkoholkonsumstörungen im Jahr 2013. Lausanne, Switzerland: Sucht Schweiz.

Marmet S, Rehm J, Gmel G, Frick H, Gmel G. 2014. Alcohol-attributable mortality in Switzerland in 2011--age-specific causes of death and impact of heavy versus non-heavy drinking. Swiss Med Wkly. 144:w13947.

Marmet S, Wicki M, Gmel G. 2017. Hospitalisierungen aufgrund von Alkohol-Intoxikation oder Alkoholabhängigkeit bei Jugendlichen und Erwachsenen - Eine Analyse der Schweizerischen „Medizinischen Statistik der Krankenhäuser" 2003 bis 2014 (Forschungsbericht Nr. 92). Lausanne: Sucht Schweiz.

Martin CS, Sher KJ, Chung T. 2011. Hazardous use should not be a diagnostic criterion for substance use disorders in DSM-5. Journal of studies on alcohol and drugs. 72(4):685-686.

Mewton L, Slade T, Teesson M. 2013. An evaluation of the proposed DSM-5 cannabis use disorder criteria using Australian national survey data. J Stud Alcohol Drugs. 74(4):614-621.

Narrow WE, Rae DS, Robins LN, Regier DA. 2002. Revised prevalence estimates of mental disorders in the United States: using a clinical significance criterion to reconcile 2 surveys' estimates. Archives of general psychiatry. 59(2):115-123.

Nelson CB, Wittchen HU. 1998. DSM-IV alcohol disorders in a general population sample of adolescents and young adults. Addiction. 93(7):1065-1077.

Pabst A, Kraus L, Piontek D, Baumeister SE. 2012. Age differences in diagnostic criteria of DSM-IV alcohol dependence among adults with similar drinking behaviour. Addiction. 107(2):331-338.

Regier DA, Kaelber CT, Rae DS, Farmer ME, Knauper B, Kessler RC, Norquist GS. 1998. Limitations of diagnostic criteria and assessment instruments for mental disorders: implications for research and policy. Archives of General Psychiatry. 55(2):109-115.

Rehm J, Room R. 2015. Cultural specificity in alcohol use disorders. Lancet (London, England).

Rohde P, Lewinsohn PM, Kahler CW, Seeley JR, Brown RA. 2001. Natural course of alcohol use disorders from adolescence to young adulthood. Journal of the American Academy of Child & Adolescent Psychiatry. 40(1):83-90.

Slade T, Teesson M, Mewton L, Memedovic S, Krueger RF. 2013. 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.

Studer J, Baggio S, Mohler-Kuo M, Dermota P, Gaume J, Bertholet N, Daeppen JB, Gmel G. 2013. Examining non-response bias in substance use research – Are late respondents proxies for non-respondents? Drug and alcohol dependence. 132(1-2):316-323.

Swiss Federal Chancellery. 2018. Alcohol in road traffic and navigation. [accessed 2018 June 25]. https://www.ch.ch/en/alcohol-in-road-traffic-and-navigation/.

Teesson M, Hall W, Slade T, Mills K, Grove R, Mewton L, Baillie A, Haber P. 2010. Prevalence and correlates of DSM-IV alcohol abuse and dependence in Australia: findings of the 2007 National Survey of Mental Health and Wellbeing. Addiction. 105(12):2085-2094.

Wakefield JC, Schmitz MF. 2014. How many people have alcohol use disorders? Using the harmful dysfunction analysis to reconcile prevalence estimates in two community surveys. Frontiers in psychiatry. 5:10.

Wakefield JC, Schmitz MF. 2015. The harmful dysfunction model of alcohol use disorder: revised criteria to improve the validity of diagnosis and prevalence estimates. Addiction. 110(6):931-942.

Winters KC, Martin CS, Chung T. 2011. Substance use disorders in DSM-V when applied to adolescents. Addiction. 106(5):882-884.

Table 1. Alcohol criteria labelling and question-wording

Criterion name	Question-wording
	In the last 12 months…
Larger/Longer	…did you often find yourself drinking more and for longer periods of time than you intended?
Hazardous use (1)	...did you more than once drive a car or another vehicle (such as a bicycle, motorcycle or moped) shortly after you had had several drinks with alcohol?
Hazardous use (2)	…did you find yourself more than once in a situation that increased your chances of getting injured (using machines, walking or doing sport in a dangerous area or around heavy traffic) after you had been drinking too much alcohol?
Tolerance	…did you find you needed a lot more alcohol to become high or drunk than you used to?
Time spent	…did you find yourself spending a great deal of time obtaining, using, or recovering from the effects of alcohol?
Neglect roles	…has your drinking alcohol caused you more than once to miss a class, work or to fail to look after your family at home?
Social/Interpersonal problems	…did you resume your drinking habits even though your drinking had caused problems with your partner, friend or acquaintances?
Withdrawal	…did you start feeling nervous or shaky for a full day or more after you had cut down on your drinking?
Quit/Control	…did you try to cut down on your drinking, but couldn't?
Craving	…have you had such a strong desire or urge to drink that you could not help drinking?
Activities given up	…did you give up activities you care about (e.g. school, work or being with friends and family) because of your drinking?
Physical/Psychological Problems	…did you continue drinking even though you were aware that alcohol had repeatedly caused you anxiety, depression or health problems?

Table 2. Descriptive characteristics at baseline (w1), first (w2) and second follow-ups (w3)

	w1	w2	w3
n	4924	4924	4924
Age (years)	19.97	21.30	25.40
SD (years)	1.22	1.26	1.24
Drinks per day			
0 drinks per day	8.9%	8.7%	8.0%
fewer than 1 per day	49.9%	49.8%	55.0%
1 per day	21.8%	21.7%	21.3%
2 per day	9.3%	10.9%	8.3%
3 or more per day	10.1%	9.0%	7.4%
Heavy episodic drinking			
6 or more drinks at least monthly	46.0%	44.1%	38.1%
maximum drinks in one day	12.15	11.46	10.27
SD	7.98	7.22	6.48
DSM-5 AUD			
mean DSM-5 AUD	1.23	1.21	1.23
SD	1.71	1.61	1.59
Prevalence of DSM-5 AUD			
mild AUD	21.3%	22.7%	23.2%
moderate AUD	6.7%	6.6%	6.5%
severe AUD	3.1%	2.5%	2.1%
any AUD (at least mild)	31.1%	31.8%	31.9%

Note: missing values for DSM-5 AUD: 12 (w1), 12 (w2), 9 (w3)

Table 3. Criterion prevalence, difficulty and discrimination, for total sample

	Prevalence (%)			Difficulty			Discrimination			Differential Item Functioning Mantel–Haenszel OR	
	w1	w2	w3	w1	w2	w3	w1	w2	w3	w1 to w2	w1 to w3
Larger/Longer	29.6	33.5	37.9	0.74	0.59	0.42	1.82	1.80	1.96	1.38	1.93
Hazardous use	27.3	29.4	32.0	1.01	0.89	0.87	1.26	1.29	1.07	1.16	1.31
Tolerance	21.7	16.4	10.4	1.11	1.43	1.80	1.71	1.62	1.71	0.57	0.26
Time spent	14.5	14.9	13.4	1.27	1.26	1.34	2.67	2.58	2.62		
Neglect roles	10.8	9.8	10.5	1.72	1.79	1.77	1.79	1.82	1.75		
Social/Interpersonal problems	4.0	4.1	3.7	2.17	2.26	2.40	2.36	2.11	1.98		
Withdrawal	3.7	3.3	3.2	2.27	2.59	2.45	2.22	1.79	2.05		
Quit/Control	3.4	3.1	4.2	2.29	2.41	2.28	2.33	2.17	2.04		1.66
Craving	2.9	2.8	3.1	2.19	2.45	2.39	2.92	2.21	2.22		1.38
Activities given up	2.7	2.2	2.6	2.21	2.40	2.40	3.01	2.69	2.41		
Physical/Psychological Problems	2.6	2.0	2.1	2.28	2.41	2.45	2.78	2.83	2.62		

Note: For Differential Item Functioning, only significant Odds Ratios (ORs with p < .05) are shown.

Table 4. Prevalence of criteria in participants with at least mild AUD and cumulative percentages for combinations of criteria

	Prevalence of criteria in participants with AUD (%)			Cumulative % of items among participants with at least mild diagnoses		
	w1	w2	w3	w1	w2	w3
n	1527	1562	1566	1527	1562	1566
Larger/Longer	73.7	78.7	84.8	0.0	0.0	0.0
Hazardous use	62.7	66.9	68.4	42.4	51.0	56.5
Tolerance	57.8	44.6	29.6	74.2	73.8	70.6
Time spent	44.0	44.3	40.4	87.0	86.2	83.3
Neglect roles	31.1	27.7	29.7	94.8	93.5	92.1
Social/Interpersonal problems	12.3	12.2	11.0	96.1	95.4	94.4
Withdrawal	11.7	9.8	9.8	97.8	97.1	95.8
Quit/Control	10.1	9.2	12.6	98.6	98.3	97.6
Craving	9.0	8.3	9.2	99.1	99.1	98.9
Activities given up	8.6	6.8	8.1	99.4	99.7	99.5
Physical/Psychological Problems	8.2	6.1	6.4	100.0	100.0	100.0

Reading example: in wave 1, 94.8% of participants with at least mild AUD would also have received that diagnosis if only the first five criteria (down to and including Neglect roles) had been included in the questionnaire.

Table 5. Prevalence of criteria in participants, with incidence in w1, w2, and w3, compared to participants with persistent AUD in waves 2 and 3

	AUD incidence in wave (%)			AUD with incidence in an earlier wave (persistent) (%)			Total with AUD in that wave (incidence + persistent) (%)			OR for difference between incident and persistent (ref.) cases			
Wave	w1	w2	w3	w1	w2	w3	w1	w2	w3	w2 unadj.	w2 adj.	w3 unadj.	w3 adj.
n	1527	660	431	0	902	1135	1527	1562	1566				
Mean age (y)	19.90	21.24	25.32		21.19	25.36	19.90	21.21	25.35				
Larger/Longer	73.7	76.7	81.9		80.1	85.9	73.7	78.7	84.8	0.82	0.97	**0.74**	0.86
Hazardous use	62.7	63.3	63.1		69.5	70.4	62.7	66.9	68.4	**0.76**	0.91	**0.72**	0.84
Tolerance	57.8	39.2	23.4		48.6	31.9	57.8	44.6	29.6	**0.68**	0.91	**0.65**	0.91
Time spent	44.0	34.0	30.2		51.8	44.3	44.0	44.3	40.4	**0.48**	**0.68**	**0.54**	0.79
Neglect roles	31.1	22.5	24.8		31.5	31.6	31.1	27.7	29.7	**0.63**	0.97	**0.72**	1.07
Social/Interpersonal problems	12.3	9.3	7.7		14.4	12.3	12.3	12.2	11.0	**0.61**	1.06	**0.59**	0.93
Withdrawal	11.7	7.9	6.5		11.2	11.0	11.7	9.8	9.8	**0.68**	1.11	**0.56**	0.88
Quit/Control	10.1	8.7	10.7		9.7	13.4	10.1	9.2	12.6	**0.89**	**1.92**	0.77	1.28
Craving	9.0	6.5	7.7		9.6	9.8	9.0	8.3	9.2	**0.65**	1.24	0.77	1.40
Activities given up	8.6	7.0	7.0		6.7	8.5	8.6	6.8	8.1	1.05	**3.06**	0.80	1.53
Physical/Psychological Problems	8.2	3.8	5.3		7.9	6.9	8.2	6.1	6.4	**0.46**	0.94	0.76	1.53
Mean AUD score/beta	3.29	2.79	2.68		3.41	3.26	3.29	3.15	3.10	**-0.204**		**-0.169**	

Note: incidence is defined as a participant's first occurrence of AUD in any of the three waves, i.e. participants with AUD and w1 and w3, but not w2, are classified as persistent in w3.
Adj.: ORs were adjusted for total score to account for the differences in AUD score between participants with incident vs persistent AUD. Given that there are only incident cases in wave 1, differences between incident and persistent cases were only tested for waves 2 and 3.
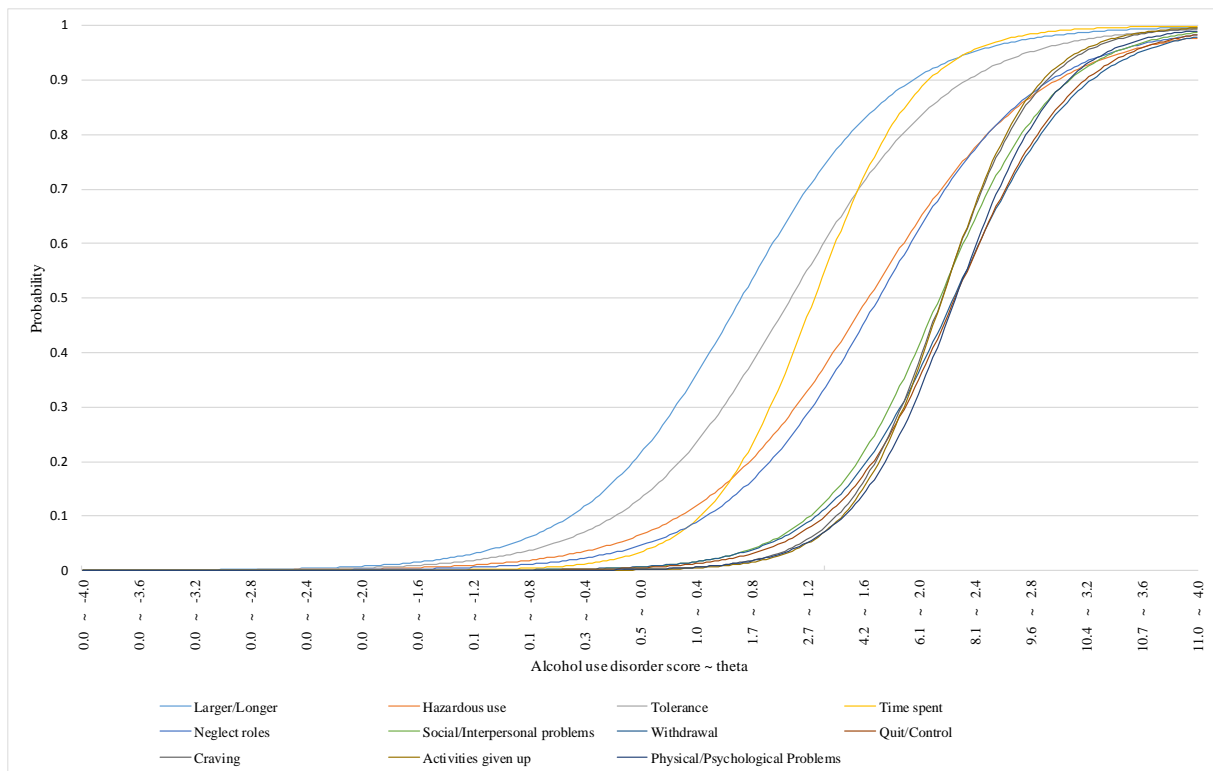
Figure 1. Item characteristic curves for the 11 alcohol use disorder criteria in wave 1

Note: The y-axis shows the probability of endorsement of a particular criterion given a certain severity of AUD measured by theta (x-axis). This corresponds to an AUD score ranging from 0–11. The difficulty of an item is the point at which it intersects the line of 0.5 probability, meaning that difficult items are to the right and easy items to the left. Items with higher discrimination have a steeper slope.
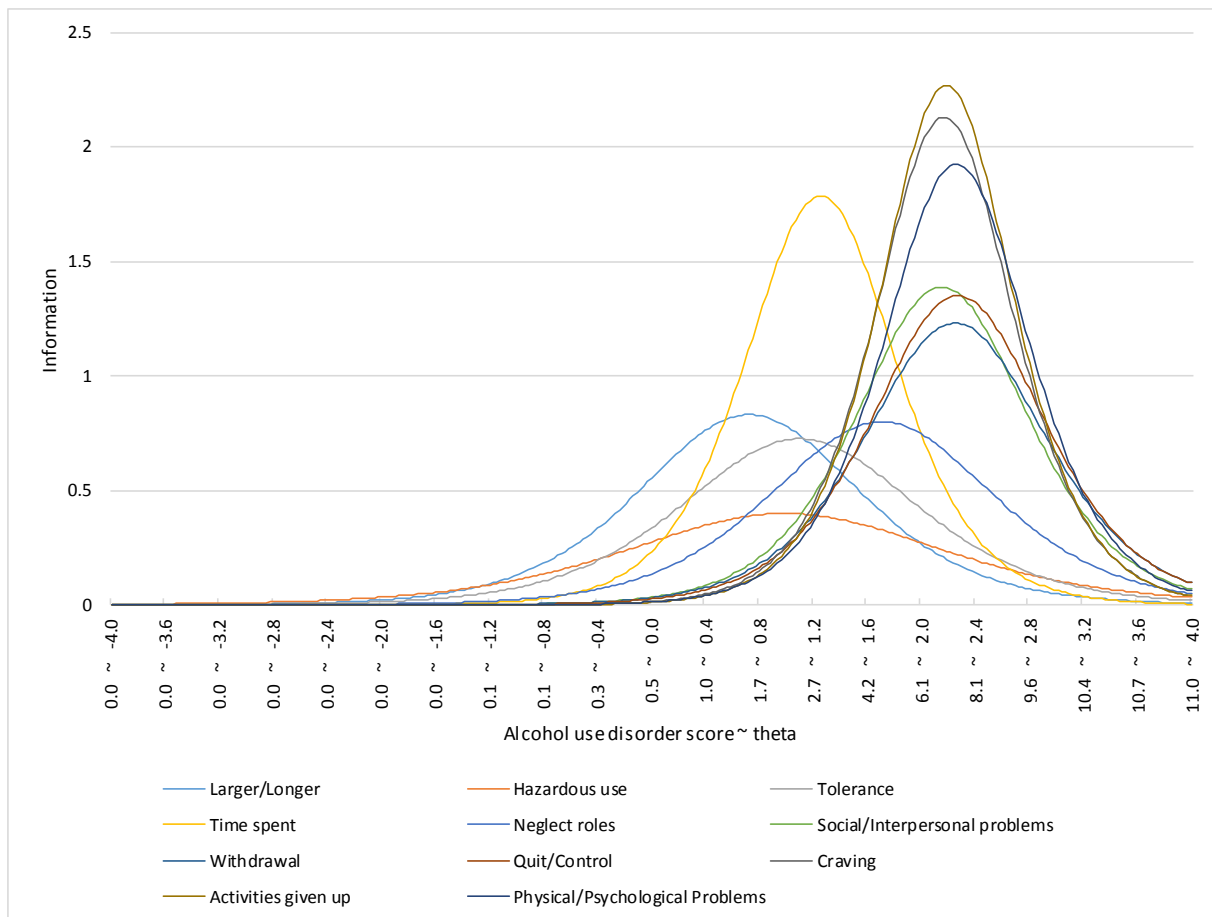
Figure 2. Item information functions for the 11 alcohol use disorder criteria in wave 1

Note: AUD=alcohol use disorder. The y-axis shows the information provided by a particular criterion given a certain severity of AUD measured by theta (x-axis). This corresponds to an AUD score ranging from 0–11. The difficulty of an item is at the highest point of its curve, meaning that difficult items are to the right and easy items to the left. Items with high discrimination provide much information at their point of difficulty, resulting in a narrower curve, whereas items with low discrimination have a broader curve, meaning that they provide information over a broader range. The reciprocal of the information function is the measurement error, meaning that where an item provides maximum information, measurement error is lowest.