Drug and Alcohol REVIEW

APSAD

Drug and Alcohol Review (March 2022), 41, 577–587
DOI: 10.1111/dar.13377

Improving measurement of harms from others' drinking: Using item-response theory to scale harms from others' heavy drinking in 10 countries

ULRIKE GRITTNER[1,2†], KIM BLOOMFIELD[1,2,3,4,5†], SANDRA KUNTSCHE[6], SARAH CALLINAN[6], OLIVER STANESBY[6] & GERHARD GMEL[7,8,9,10]

[1]*Institute of Biometry and Clinical Epidemiology, Charité – Universitätsmedizin Berlin, Berlin, Germany,* [2]*Berlin Institute of Health, Berlin, Germany,* [3]*Centre for Alcohol and Drug Research, Aarhus University, Copenhagen, Denmark,* [4]*Health Promotion Department of Public Health, University of Southern Denmark, Esbjerg, Denmark,* [5]*Alcohol Research Group, Public Health Institute, Emeryville, USA,* [6]*Centre for Alcohol Policy Research, School of Psychology and Public Health, La Trobe University, Melbourne, Australia,* [7]*Alcohol Treatment Centre, Lausanne University Hospital, Lausanne, Switzerland,* [8]*Research Department, Addiction Switzerland, Lausanne, Switzerland,* [9]*Centre for Addiction and Mental Health, Institute for Mental Health Policy Research, Toronto, Canada, and* [10]*Faculty of Health and Applied Science, University of the West of England, Bristol, UK*

Abstract

Introduction. *The heavy drinking of others may negatively affect an individual on several dimensions of life. Until now, there is scarce research about how to judge the severity of various experiences of such harms. This study aims to empirically scale the severity of such harm items and to determine who is at most risk of these harms.* ***Methods.*** *We used population-based survey data from 10 countries of the GENAHTO project (Gender and Alcohol's Harms to Others, data collection: 2011–2016). Questions about harms from others' drinking asked about verbal and physical harm, damage of belongings, traffic accidents, harassment, threatening behaviour, family and financial problems. We used item response theory methods (IRT) to scale severity of the aforementioned items. To acknowledge culturally based variations in different countries, we assessed 'differential item functioning'.* ***Results.*** *The items 'family problems', 'financial problems' and 'clothes and property damage' as well as 'physical harm' were scaled as more severe in most countries compared to other items. Substantial differential item functioning was present in more than half of the country pairings. The item 'financial problems' was most often differentially scaled. Younger people who drank more, as well as women (compared to men), reported more harm.* ***Discussion and Conclusions.*** *Using IRT, we were able to evaluate grades of severity in harms from others' drinking. IRT scaling yielded in similar rankings of items as reported from other studies. However, empirical scaling allows for more differentiated severity scaling than simple summary scores and is more sensitive to cultural differences.* [Grittner U, Bloomfield K, Kuntsche S, Callinan S, Stanesby O, Gmel G. Improving measurement of harms from others' drinking: Using item-response theory to scale harms from others' heavy drinking in 10 countries. *Drug Alcohol Rev* 2022;41:577–587]

Key words: alcohol, harms to others, scaling, item response theory.

Introduction

Varying approaches have been proposed to measure alcohol-related problems that are regarded as harm that the drinker causes to other people. No standardised instrument has yet been developed to quantify alcohol harm that is experienced by others.

Research on alcohol's harms to others (AHTO) has grown substantially in recent years [1], the wide spectrum of AHTO includes harms at the societal level, such as economic costs to society [2], traffic accidents [3], violence and physical abuse [4], foetal alcohol spectrum disorder [5] and crime, as well as public order and safety [6]. Another area of harm comes from

†Ulrike Grittner and Kim Bloomfield should be considered joint first authors.

Ulrike Grittner PhD, Research Associate, Kim Bloomfield DrPH, Professor, Sandra Kuntsche PhD, Associate Professor, Sarah Callinan PhD, Senior Research Fellow, Oliver Stanesby MSc, Research Officer, Gerhard Gmel PhD, Senior Research Fellow. Correspondence to: Dr Ulrike Grittner PhD, Research Associate, Institute of Biometry and Clinical Epidemiology, Charité Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany. Tel: 0049 30 450 562 174; Fax: 0049 30 450 562 972; E-mail: ulrike.grittner@charite.de

Received 8 September 2020; accepted for publication 14 July 2021.

official police registries on drinkers' harms to the family (including child abuse) [7,8]. Harms that are reported at the individual level have included those caused by the drinking of friends, acquaintances [9–11], work colleagues [12] and strangers [13].

Information on societal harm and reported family abuse are usually collected as official statistics. Harms committed as verbal insults by drunken friends or strangers, absenteeism of colleagues at the workplace, or child abuse as witnessed by family members are often reported in monitoring questionnaires. Sometimes these questionnaires ask respondents how much the harm had affected them [10,14].

Several researchers have stressed that it is not sufficient to use the sum of harm items, within which all items presumably carry equal importance. Therefore, there is the need to *quantify the severity* of individual AHTO items [10,15–18]. Such a metric has yet to be incorporated into the measurement of AHTO items.

An initial step was taken by Callinan and Room [14] in which they grouped AHTO items based on a survey of young adults in Victoria, Australia. Multiple correspondence analysis identified response patterns. Two groups emerged: 'amenity harms' and 'tangible harms'. The latter group contained items of greater severity, while the former group contained items that 'expressed fears and loss of amenity'. In another study, Stanesby *et al.* [1] gathered data from 36 key alcohol research and policy experts from 23 countries who ranked 48 AHTO items by severity of perceived harm. Key informants rated physical, financial, practical and severe emotional harms as most severe, while harms involving verbal insults and annoyances were rated as least severe. The strength of this study was that it attempted to rate many harms along a severity scale. However, the experts commented that rating severity was difficult, since the harm experience was not described in a specific setting, the gender of victim and perpetrator was not given, and because the nature of the victim-perpetrator-relationship was unclear [1]. So, although these studies have begun to attempt a ranking of AHTO items, there remains a need to base rankings on the assessments of the general population. The present paper quantifies the severity of harms from others' drinking on the basis of representative general population samples in 10 countries.

Any attempt to quantify the severity of AHTO for international use first must acknowledge the challenge in trying to develop a standardised scale that takes into account varying drinking cultures and perceptions of alcohol harm. Researchers have experienced similar challenges before [19–21], as in the Cross-Cultural Applicability Study which spent many years studying the meanings of diagnostic criteria for alcohol dependence and harmful use in various countries [19]. The study revealed that translations and understandings of dependence criteria such as intoxication and withdrawal varied greatly across countries or did not even exist in some. Room *et al.* [19] describe how subjects in their nine study countries perceived the severity of dependence criteria very differently. For example, the authors describe a man in Bangalore, India who reported that he drank never more than two standard bottles of beer once every 2 months, that his drinking had less effect on him than earlier, that his family and friends objected to his drinking, and that his doctor had recommended that he stopped his drinking. This was enough to qualify for a diagnosis of dependence according to the Composite International Diagnostic Interview criteria, which illustrates the challenge that 'in a cultural situation where there is much disapproval of drinking, the threshold for positive responses to precoded questions has been set very low, so that a mechanical application of scoring algorithms for dependence would result in inappropriate diagnosis' [19, p. 217].

In the present study, we have attempted to scale a common set of AHTO items based on general population surveys. Using data from a variety of countries (four European, one South American and five Asian) brings together various drinking cultures. As mentioned, what might be seen as a severe harm item in one country, might be seen as less harmful in another. Therefore, our task was not only to find common ratings of harm items, but also to assess what differences in scaling arise across countries. Items that are scaled differently may reveal cultural variations in drinking norms and habits. Additionally, we aimed to demonstrate how the scaled harm items can be used for future research to identify groups most vulnerable to AHTO.

Thus, the research questions of the present study were:

1. Is it possible to develop a common scale to measure severity of AHTO items across 10 countries? Or is it more appropriate to use different scales for groups of countries (or individual countries)? What are similarities and differences in the scaling between countries?
2. How can groups of respondents who are more affected by AHTO than other groups be characterised?

Methods

The data come from the GENAHTO project, which has collected cross-sectional surveys from several countries (Table 1). Data collection took place between 2011 and 2016 and sample sizes range between 1007 and 4813 per country. In three countries, telephone surveys (one

Table 1. *Survey characteristics of 10 GENAHTO study countries*

Country	Survey year	Administration mode	Response rate	Sampling frame	*n*	Cases with data on 10 items
Switzerland	2012/2016	Tel (CATI)	51%/45%	National	4892	4813
Denmark	2011	Web/tel (CATI)	64%	National	5133	2545[a]
Ireland	2015	Tel (CATI)	37%	National	2005	1991
Scotland	2012	Face-to-face	Unknown	National	1007	1007
Chile	2012/2013	Face-to-face	72%	Regional	1500	1364
India	2013/2014	Face-to-face	97%	Regional	3403	3356
Sri Lanka	2013/2014	Face-to-face	93%	National	2475	2475
Thailand	2012/2013	Face-to-face	94%	National	1695	1695
Vietnam	2012/2013	Face-to-face	99%	National	1501	1470
Lao PDR	2013	Face-to-face	99%	National	1257	1252
Total					**24 868**	**21 968**

[a]Some questions were only asked in a random sub-sample of 2569 respondents. CATI, computer assisted telephone interview; GENAHTO, Gender and Alcohol's Harms to Others; Tel, telephone; Web, online questionnaire.

country combined web-based and telephone surveys) were conducted; in seven countries, face-to-face interviews were conducted. Country surveys were approved by each respective Institutional Review Board. The overall project was approved by the Institutional Review Board of the Public Health Institute in the USA. The GENAHTO project is described in detail in Wilsnack *et al.* [22,23]. We used information only from participants who had complete data on all 10 harm items.

Ten AHTO items were asked of respondents. They were posed as yes–no questions: 'In the last 12 months has someone who had been drinking':

1. *called you names*, or otherwise *insulted* you?;
2. *harassed or bothered you/made you feel threatened or afraid* at home or in some other private setting?;
3. *ruined your clothes or other belongings?/damaged your house, car or property*?;
4. *harmed you physically*?;
5. been responsible *for a traffic accident* you were involved in?
6. been responsible *for family problems* or *marriage difficulties*?;
7. been responsible for *financial trouble*?;
8. *harassed or bothered you on the street* or in some public place?;
9. made you *afraid on the street*?;
10. been making you been *kept awake at night* by drunken noise?

Statistical methods

Scalability of items across countries. Using exploratory factor analysis (with tetra-choric correlation and two un-rotated factors), we first evaluated whether items are scalable on one dimension. Using a confirmatory factor analysis with a one-factor solution, we determined whether all items loaded sufficiently on one dimension (loading criterion: >0.3). Cronbach's α was calculated as a measure of internal consistency (range: 0.54–0.82), while Mokken's H was examined as a coefficient of homogeneity (sufficient unidimensionality: Mokken's $H > 0.3$ for seven countries, Mokken's H between 0.2 and 0.3 for three countries; Table S2, Supporting information) [24].

Country-specific item response theory methods analyses. Analyses within an item response theory (IRT) paradigm are predicated on the assumption that both respondents and items can be measured on the same scale [25]. In IRT, the participant characteristic of interest (here: harm affectedness by the drinking of others) is seen as a latent trait, which is measured by using various binary items. The probability that a participant experienced one of these harm items is determined by the *severity* of the item (more severe items are experienced by fewer respondents) and by the *affectedness* level of the subject (latent trait). The so-called item characteristic function of an item is a monotone probability function. IRT models are widely used in educational testing since they have high measurement accuracy and reliability, thereby allowing the use of flexible sets of items for different subjects, for example, via computerised adaptive testing. IRT has become more popular in recent years in the health sciences, especially in quality of life research and for scaling functional outcomes in patients [26]. Since IRT can be used to scale different items based on the response patterns of participants, this method is also suitable to empirically quantify the severity of AHTO

items. The use of IRT allows for: (i) scaling of *items* with regard to AHTO; and (ii) scaling of *persons* on AHTO. With this approach, it is possible to quantify the degree of severity of the experienced harm of each item. The scaling of the items is based on the probability of endorsement: low scaling values represent less severe items endorsed by many respondents, whereas high scaling values represent more severe items endorsed by fewer respondents. Items are scaled as more harmful if individuals who experienced it also experienced more often situations described in less harmful items.

Additionally, in two-parameter logistic (2PL) IRT models, it is possible to estimate discrimination properties (slopes) of the items, in contrast to classic Rasch models that have only one parameter (location: harm severity). Low discrimination items have a limited ability to differentiate between highly and less affected people. This happens only if fewer people endorse an item, and when only some of those have endorsed other less severe items, implying that 'lighter' items do not affect them.

The 2PL-IRT models were calculated separately for each country. Parameters for severity and discrimination with 95% confidence intervals are reported. For comparisons of model fit, we also calculated one-parameter models (Rasch models, 'ltm' package of R [27]) and more complex three-parameter models which additionally estimate specific item-intercepts. Using likelihood-ratio tests, we compared the different models for each country in order to determine whether the 2PL models gave the best model fit.

Differential item functioning. To acknowledge culturally based variations/sensitivities to AHTO, we used pair-wise differential item functioning (DIF) analyses (criterion: relative beta change of >0.10, in R package 'lordif') [28]. We therefore tested group effects in binary logistic models for the probability of experiencing a specific harm, simultaneously adjusting for the latent variable level (AHTO).

We used a 'beta change' criterion, meaning that the relative change of the regression coefficient for AHTO between two models (with and without country as covariate) should be at least 10% to identify DIF. This definition models uniform DIF and accounts for differences in the severity, but not in discrimination estimates. Using an iterative process, the smallest number of items with DIF between two countries was identified. If DIF is present, that means that the described experience is seen as more harmful in one country compared to another country.

'Affectedness' score of AHTO. For each participant, we calculated a score value for being affected by AHTO based on country-specific scalings. To answer the question of which characteristics were related to levels of being affected by AHTO, we analysed these scores in multiple linear mixed models (random intercept for country) with regard to administration mode (telephone/web or face-to-face), sex, age, education and drinking level (average intake per day, log-transformed).

We report two multiple regression models: one with main effects only, and a second with additional interaction effects. The final model was chosen on the basis of model comparisons with all possible two-way interactions using the Bayesian information criterion. Percentages and regression estimates are based on weighted data to account for participants' probability of being selected to participate (according to the number of eligible persons present in the household) and non-response (to match the gender distribution in each country), and to match the national adult population distributions of age.

Results

Response rates were very high in five of the 10 countries and ranged between 93% and 99% (India, Sri Lanka, Thailand, Vietnam, Lao PDR; Table 1). However, in Chile the response rate was 72%, in Denmark 64%, in Switzerland 51% in 2012 and 45% in 2016 and in Scotland the response rate is unknown. Overall rate of cases with missing data (item non-response) was low with 1.5%, with the exception of Chile with a missing value rate of 9.1%. In light of this low rate of cases with missing data, the risk of bias is negligible. Only cases with complete data were used in this study.

Unidimensionality and scalability

In total, we used information from 21 968 participants in 10 countries (Table 1). Table 2 displays the prevalence of harms by country. While the prevalence for 'awake at night' (25.6% for total), 'insult' (21.9%), 'harassment on the street' (18.2%) and 'afraid on the street' (18.2%) were high in most of the countries, items such as 'financial problems' (6.3%), 'physical harm' (4.6%) and 'traffic accident' (2.6%) had a low prevalence in most countries.

Exploratory factor analysis across all countries with un-rotated factors (two-factor solution) resulted in item loadings of 0.55 or higher on the first factor and explained 49% of the total variance. A one-factor (confirmatory) solution produced factor loadings of 0.4 or higher on each item. The *H*-Index in the confirmatory factor analysis was 0.3 or higher for each item, Mokken's *H* for all items was 0.35, and Cronbach's α

Table 2. *Prevalence of harm items by country (%), ordered by total prevalence (percentages based on weighted data, unweighted* n*)*

Country	*n*	Awake at night	Insult	Harassment on the street	Was afraid on the street	Harassment/ threatening at private place	Family problems	Damage of clothes property	Financial problems	Physical harm	Traffic accident
Switzerland	4813	22.3	10.9	18.1	12.5	10.0	2.3	4.1	0.1	1.2	0.3
Denmark	2545	26.2	21.0	25.9	14.0	13.6	6.3	4.9	0.8	2.5	0.8
Ireland	1991	26.0	21.4	22.5	15.8	21.0	10.7	14.8	2.2	2.1	0.6
Scotland	1007	30.2	18.9	20.2	16.4	15.5	5.6	7.0	1.5	4.9	1.2
Chile	1364	34.3	27.4	31.7	22.6	23.3	11.1	11.5	5.3	6.5	3.2
India	3356	35.5	29.0	20.1	38.0	23.5	16.1	11.0	15.2	12.5	5.3
Sri Lanka	2475	16.7	27.9	10.6	10.9	13.9	7.1	10.1	15.2	7.4	6.1
Thailand	1695	32.1	30.0	15.3	31.8	22.4	8.6	8.6	11.6	2.5	5.5
Vietnam	1470	11.2	26.8	12.8	4.7	23.7	8.1	7.1	5.6	6.4	2.8
Laos	1252	20.4	16.9	6.8	17.2	10.6	3.5	3.6	5.8	2.4	3.7
Total	**21 968**	**25.6**	**21.9**	**19.4**	**18.2**	**16.6**	**7.8**	**7.7**	**6.3**	**4.6**	**2.6**

was 0.73. All of these characteristics demonstrated sufficient scalability of all items.

Country-wise scaling

Comparisons of one-, two- and three-parameter models demonstrated the best model fit of 2PL models for each country. Therefore, we report only results of the 2PL solutions. For most countries, the items 'insult', 'harassment/threatening at a private place', 'awake at night', 'harassment on the street' and 'afraid on the street' were scaled as 'less severe' (Figure 1). In contrast, 'family problems', 'financial problems', 'clothes or property damage' and 'physical harm' were scaled as more severe items. For all countries, 'traffic accidents' was scaled as the most severe item. However, level of severity and ordering of items differed across countries.

As an example, we compare the scaling results in Switzerland and Denmark (Figure 1, Table 3). Both countries had similar item scalings and no DIF (Table 3), meaning that all items were scaled in a similar way, with 'traffic accidents' as the most severe harm. However, severity scores for 'awake at night' differed between Denmark (1.6) and Switzerland (3.7) (Figure 1), but this item had a low discrimination score in both countries, meaning that it did not discriminate well between severely affected and less severely affected people for most countries, and especially so in Switzerland.

Differential item functioning

Table 3 displays the results of the DIF analysis, which indicates which items cannot be regarded as similarly valued items across the respective countries. Substantial DIF was detected for one or more AHTO items in 26 different pairings of countries (out of 45 possible pairs of countries). There were 10 country pairings with DIF for only one item. Countries which had several DIF item pairings were Sri Lanka, Laos (DIF items in seven different country pairings) and Vietnam (DIF in six country pairings). Overall, we found for many pairings of countries items that scaled with DIF (Table 3; Table S1, Supporting information). Additional sensitivity analyses of potential DIF in each country by sex and age (<40 years, ≥40 years) did not reveal any major DIF apart DIF by age for 'financial problems' and 'awake at night' in Switzerland and DIF by age for 'traffic accident' in Ireland. However, since sample sizes in subgroups were low, these differences were not further investigated.

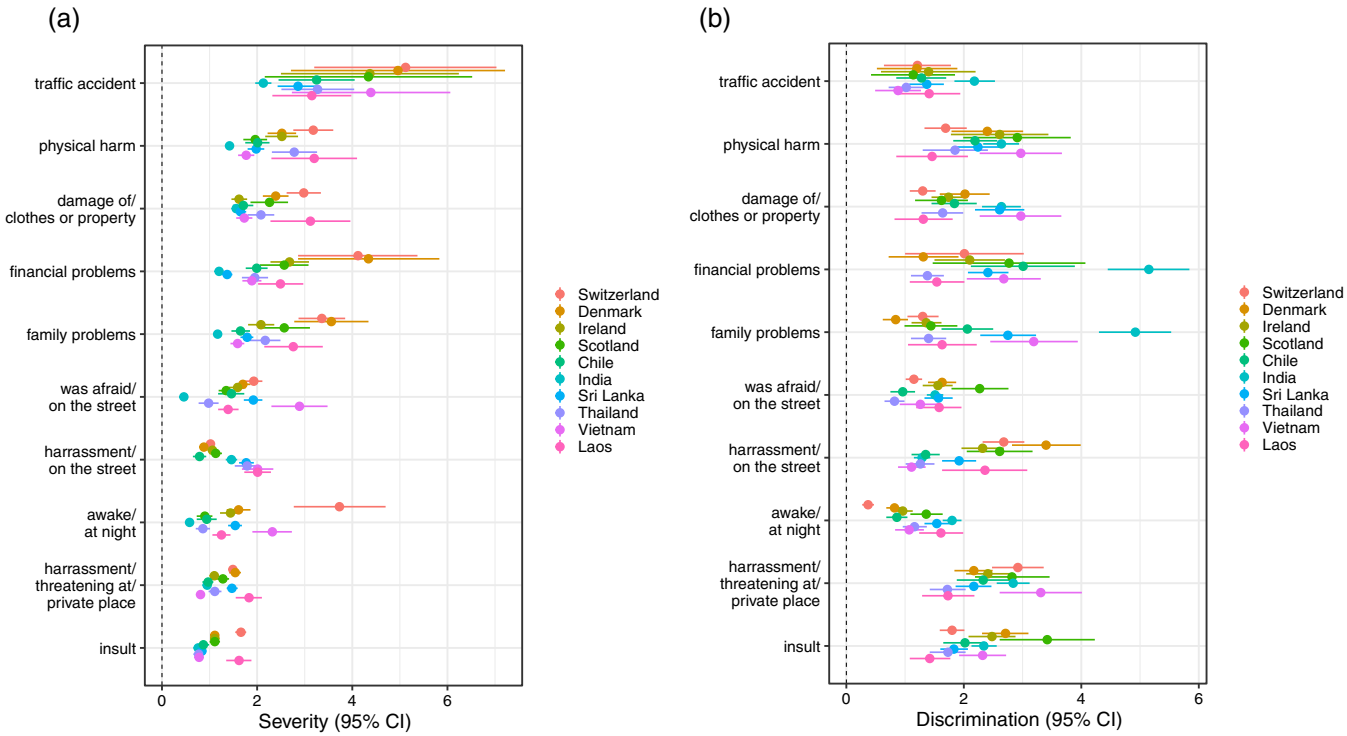

Figure 1. *Severity (a) and discrimination (b) estimates of harm items by country ordered by total severity scores. CI, confidence interval.*

Table 3. *Pairwise differential item functioning, criterion beta change of 0.1 or more*

	SWI	DEN	IRE	SCO	CHIL	IND	SRI	THAI	VIET	LAO
SWI		–	1	–	1	3	5	2	5	1
DEN			–	–	–	1	1	3	4	2
IRE				–	–	4	4	–	–	4
SCO					–	–	1	–	2	2
CHIL						–	2	1	–	2
IND							3	–	1	–
SRI								–	–	1
THAI									1	–
VIET										3
LAO										

–: no items with DIF; 1: one item with DIF; 2: two items with DIF; 3: three items with DIF; 4: four items with DIF; 5: five items with DIF. Twenty-six pairs with DIF out of 45 pairings, 10 times 1 item with DIF, 6 times 2 items with DIF, 4 times 3 items with DIF, 4 times 4 items with DIF, 2 times 5 items with DIF. CHIL, Chile; DEN, Denmark; DIF, differential item functioning; IND, India; IRE, Ireland; LAO, Laos PDR; SCO, Scotland; SRI, Sri Lanka; SWI, Switzerland; THAI, Thailand; VIET, Vietnam.

Table 4. *Linear mixed model, outcome: theta (*100 rescaled from IRT scaling), random intercept for country (coefficients based on weighted data)*

	Only main effects (M0)		With interactions (M1)	
	Marginal R^2: 0.07, conditional R^2: 0.07		Marginal R^2: 0.09, conditional R^2: 0.09	
Variables	β (95% CI)	*P* value	β (95% CI)	*P* value
Fixed effects				
Intercept	44.7 (40.7–48.7)	<0.001	48.3 (44.5–52.1)	<0.001
Administration mode telephone or web (ref: face-to-face)	−6.6 (−12.3 to −1.0)	0.047	−9.5 (−14.2 to −4.9)	0.003
Sex (females)	3.0 (2.9–3.8)	0.011	3.0 (0.8–5.2)	0.009
Age (years)	−1.12 (−1.19 to −1.10)	<0.001	−1.15 (−1.22 to −1.09)	<0.001
Education (ref: <high school)				
At least some higher education	−10.6 (−13.4 to −7.9)	<0.001	−7.6 (−10.3 to −4.8)	<0.001
Completed high school	−8.8 (−11.6 to −6.1)	<0.001	−6.3 (−9.0 to −3.6)	<0.001
Volume (in g day^{-1}) (log-transformed)	6.5 (5.8–7.3)	<0.001	16.4 (15.0–17.7)	<0.001
Volume (in g day^{-1})*2 (log-transformed)			−0.7 (−1.1 to −0.3)	
Interaction age*volume (log-transformed)			−0.28 (−0.32 to −0.25)	<0.001
Interaction sex (females)*volume (log-transformed)			−1.9 (−3.0 to −0.8)	0.001
Random effects (SD)				
Intercept (country)	3.4 (1.5–5.6)		2.6 (0.7–4.4)	
Residual	75.2 (74.5–76.0)		74.6 (73.9–75.4)	

Independent variables: administration mode (telephone/web or face-to-face), age, sex, education, volume (log-transformed), interactions between: education*age, volume*age, sex*volume, volume*education, n = 20 610 individuals in nine countries (models compared according to Bayesian information criterion). CI, confidence interval.

Who is more/less affected by AHTO?

Table 4 displays the regression coefficients of scores for being affected by AHTO, which are based on multiple linear mixed models. Women had higher scores than men. Younger respondents were more affected than older respondents were. Those with lower educational achievement were more affected by the negative consequences of the alcohol consumption of others than those with middle or higher education. Own drinking level was positively associated with the score.

In the more complex model, a quadratic term for alcohol volume (log-transformed) was included to reflect the curvilinear positive relationship between volume and the AHTO score, with a steep slope for none to low or middle alcohol volume and a plateau for higher volume levels (Table 4, Figure 2). There was an interaction between *age* and *volume* such that own volume of drinking was strongly and positively associated with the 'being affected' score among young respondents. For the oldest respondents, there was a weak negative association between own alcohol consumption (volume) and score levels. The interaction between *sex* and *volume* indicates that sex differences in score levels with higher scores for women were especially present for those with none or low alcohol consumption. In contrast, among respondents with high alcohol consumption, men had higher scores than did women.

Discussion

In this study, we have shown that modelling based on the 2PL-IRT approach functioned appropriately for scaling AHTO items. For all study countries, the item 'traffic accidents' scaled as the most severe item. Furthermore, for most countries 'family problems', 'financial problems', 'clothes or property damage' and 'physical harm' were scaled as more severe items, whereas the items 'insult', 'harassment/threatening at a private place', 'awake at night', 'harassment on the street' and 'afraid on the street' were scaled as less severe.

In the growing AHTO literature, it is known that items that were rated here as less severe represent 'nuisance' items [29], revealing a crude but common-sense evaluation of these items. Our analyses largely agree with the results of a recent study [1]. The overlapping agreement between these two studies can be seen as

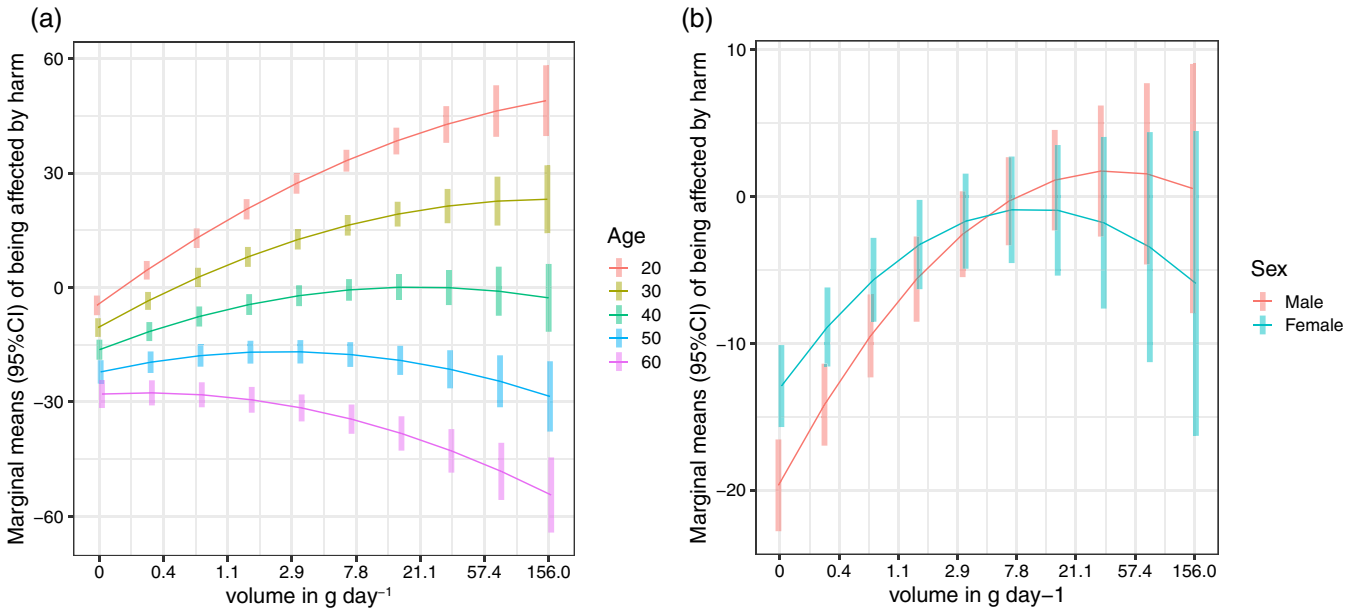

Figure 2. *Adjusted marginal fixed estimates of linear mixed models for the interaction of age*volume (a) and sex*volume (b).*

triangulation in the validation of AHTO scaling approaches [30]. It should be noted that both studies included lower income societies [1].

Although the rank ordering of the AHTO items was similar across most countries (as depicted in Figure 1), DIF was found for at least one AHTO item in more than half of all country-wise pairings. Differing cultural norms and habits might have varying levels of influence to these differential weights of severity. The presence of DIF in many pairings indicates that it is impossible to construct a single scale of AHTO items that would be valid across the study countries.

The items 'financial problems', 'traffic accidents' and 'harassment on the street' showed most often DIF. However, traffic accidents were overall rarely experienced and the item was scaled as most severe in all countries except Laos. In contrast to the item 'harassment on the street', the item 'financial problems' was scaled on very different levels of severity across countries (Figure 1; Table S3, Supporting Information). This is not surprising given the wide spectrum of countries in our study. The low- to (upper) middle-income countries included India, Laos, Sri Lanka, Thailand and Vietnam, while the high-income countries were Chile, Denmark, Ireland, Scotland (UK) and Switzerland [31]. In 2017, the adjusted mean net national income was approximately $41 800 in the high-income countries and $2990 in the low- to middle-income countries of our study [32]. Based on 2017 beer pricing data [33], the ratios of price-to-income for our study countries were 1.7×10^{-4} for the high-income countries and 8.8×10^{-4} for low-to-middle income countries, meaning that alcohol in the low- and middle-income countries is over five times more expensive than in the high-income countries. Purchasing alcohol can therefore have a direct economic impact on persons and families in lower and middle-income countries [34,36]. Thus, the economics of drinking in the various countries may have led to inconsistencies in scaling. 'Financial problems' are more common in lower-to-middle income countries but are scaled as less severe due to the fact that they are more common as compared to higher income countries. Thus, the affordability argument is only one possibility of how alcohol is linked to financial harms.

Another explanation is one of how much importance 'financial problems' represents across low-to-high income countries. One could argue that reporting financial harms due to someone's drinking may signal something more 'extreme' in high-income countries than in low-income countries. The reasoning would be that the threshold over which financial harm would occur due to someone else's drinking would be higher in high-income countries than in low-income countries. The argument is that in high-income countries financial problems due to another person's drinking might be related to a collection of more threatening problems, such as job loss, difficulties to pay rent, mortgage or a loan or costs of divorce. These problems seriously affect the respondent with regard to various dimensions, such as personal (mental, emotional), business and legal relationships. Of course, these extreme problems also occur in low-income countries, but we assume that in low-income countries more respondents who report financial problems due to someone's drinking will also include the reporting of the direct

effects of buying expensive alcohol [36], while in high-income countries, this may not be the case.

Despite several sensitivity analyses by excluding one item and excluding countries, it was not possible to find a larger group of countries and items for robust cross-country scaling. In addition to the absence of a common scale across countries, it is, however, possible to use the country-specific scaling results in national or international studies as a quantitative measure of investigating how much someone is affected negatively by others' drinking. The country-specific standardised θ values of severity for each item can easily be used as item scores and can be summed up across different items (Table S3). We have illustrated the use of scaled items in our analysis of the association of participant's characteristics to the score of being affected by harm of others' drinking and demonstrated in a cross-country analysis that younger age, female sex and higher levels of own drinking is associated to a higher load of negative consequences due to someone else's drinking (Table 4, Figure 2).

In examining who was more likely to report AHTO, our analyses confirmed the work of Laslett *et al.* [36] and demonstrated that younger people reported more severe harm than older people, women experienced more harm than men, own drinking levels were positively associated with AHTO levels and low educational level was associated to higher levels of AHTO. However, we also discovered some new insights, such as a positive curvilinear relationship between own drinking level and AHTO, with a ceiling effect achieved at higher drinking levels. We also found a steep positive relationship between own drinking level and AHTO in younger people (below 40 years) but not in older people. These differences might be related to the different life circumstances of younger people and to their particular drinking environments, such as drinking at parties, bars or group events. In agreement with these findings, Bond *et al.* [36] showed in an international study that younger people drink more frequently in public venues compared to older people while this age effect is not present for drinking in private venues. Several studies have additionally shown that young people experience harm in public drinking settings [37,38], and Bahler and Sundaram [39] and Marmet and Gmel [40] have shown for Switzerland, that more negative consequences occur in public drinking settings. Finally, at very high levels of drinking men reported more AHTO than did women.

Limitations

Among the limitations of our study was the fact that we could not include more countries and more harms, as not all survey questionnaires were comparable. Additionally, it would have been better to use more specific harms that also included information on the relationship of the respondent to the drinker (stranger/known drinker). Again, this was not possible due to the lack of comparable questions across countries. Furthermore, some questions that were included were highly, but not completely, comparable in wording across countries. Response rates of study countries varied widely with higher rates in low- to middle-income countries where face-to-face interviews were conducted (Table 1). We cannot exclude potential bias due to low response rates in some countries and we might have especially missed respondents who were more strongly affected with regard to AHTO. Therefore, more representative surveys might end up with a different scaling, in particular with other quantifications of the items. However, we think that the ordering of the items would be comparable. As for Chile and India, only regional samples were available, the results might have been different if national samples would have been available for those two countries. Because of limited sample sizes, we could not pursue gender-specific scales of harm items. This will have to remain a task for future large studies.

Implications

Our study provides an empirically based quantitative assessment of the severity of AHTO items from an international perspective. Scaling of AHTO items is helpful in gaining insight into differential weighting of items with regard to harm severity. Our results provide indications to policymakers as to which harm items are seen as most grave and to what extent these perceptions are held in common across countries. Additionally, our identification of groups most vulnerable to AHTO is more 'precise' when using the IRT-scaled values than when merely using summary measures of items in which nuisance harms and more severe harms carry the same weight. The fact that we have been able to find some commonly experienced harms as well as to have identified risk groups provides more 'clout' for tackling AHTO internationally [41]. Thus, various agencies now have a growing knowledge base for developing both prevention and assistance programs for those who suffer most from AHTO.

Acknowledgements

The data used in this paper are from the GENAHTO Project (Gender and Alcohol's Harm to Others),

supported by NIAAA Grant No. R01 AA023870 (Alcohol's Harm to Others: Multinational Cultural Contexts and Policy Implications). GENAHTO is a collaborative international project affiliated with the Kettil Bruun Society for Social and Epidemiological Research on Alcohol and coordinated by research partners from the Alcohol Research Group/Public Health Institute (USA), University of North Dakota (USA), Aarhus University (Denmark), the Centre for Addiction and Mental Health (Canada), the Centre for Alcohol Policy Research at La Trobe University (Australia) and the Addiction Switzerland Research Institute (Switzerland). Support for aspects of the project has come from the World Health Organization, the European Commission (Concerted Action QLG4-CT-2001-0196), the Pan American Health Organization, the Thai Health Promotion Foundation, the Australian National Health and Medical Research Council (Grant No. 1065610) and the US National Institute on Alcohol Abuse and Alcoholism/National Institutes of Health (Grants R21 AA012941, R01 AA015775, R01 AA022791, R01 AA023870 and P50 AA005595). Support for individual country surveys was provided by government agencies and other national sources. National funds also contributed to collection of all of the datasets included in World Health Organization projects. Study directors for the survey datasets used in this paper have reviewed the paper in terms of the project's objective and the accuracy and representation of their contributed data. The study directors and funding sources for data sets used in this report are: *Switzerland*—Gerhard Gmel, Federal Office of Public Health (Grant No. 13.007278/204.0001/ – 1199); *Denmark*—Kim Bloomfield, Centre for Alcohol and Drug Research, Business and Social Sciences, Aarhus University; *Ireland/Scotland*—Ann Hope, Health Service Executive, Department of Health, National Office for Suicide Prevention and the National Office for the Prevention of Domestic, Sexual and Gender-based Violence (COSC), Department of Justice; *Chile*—Ramon Florenzano, Thai Health Promotion Foundation, World Health Organization; *India*—Vivek Benegal and Girish Rao, Thai Health Promotion Foundation, World Health Organization; *Sri Lanka*—Siri Hettige, Thai Health Promotion Foundation, World Health Organization; *Thailand*—Orratai Waleewong and Jintana Janchotkaew, Thai Health Promotion Foundation, World Health Organization; *Vietnam*—Hanh T. M. Hoang and Hanh T. M. Vu, Thai Health Promotion Foundation, World Health Organization; *Lao PDR*—Latsamy Siengsounthe, Thai Health Promotion Foundation, World Health Organization.

References

[1] Stanesby O, Gmel G, Graham K, Greenfield TK, Waleewong O, Wilsnack SC. Improving measurement of harms from others' drinking: a key informant study on type and severity of harm. Nordisk Alkohol Nark 2020;37:122–40.

[2] Kjellberg J, Poulsen CA. Samfundsøkonomiske omkostninger forbundet med alkoholoverforbrug [Socio-economic costs associated with heavy alcohol use]. Copenhagen: The Danish Centre for Social Science Research, 2014.

[3] Connor J, Casswell S. The burden of road trauma due to other people's drinking. Accid Anal Prev 2009;41:1099–103.

[4] McClelland GM, Teplin LA. Alcohol intoxication and violent crime: implications for public health policy. Am J Addict 2001;10:70–85.

[5] Popova S, Lange S, Probst C, Gmel G, Rehm J. Estimation of national, regional, and global prevalence of alcohol use during pregnancy and fetal alcohol syndrome: a systematic review and meta-analysis. Lancet Glob Health 2017;5:e290–9.

[6] Klingemann H. Public order and safety. In: Klingemann H, Gmel G, eds. Mapping the social consequences of alcohol consumption. Dordrecht: Kluwer Academic Publishers, 2001:113–32.

[7] Livingston M. Effects of alcohol consumption in spousal relationships on health-related quality of life and life satisfaction. J Stud Alcohol Drugs 2009;70:383–90.

[8] Lown EA, Nayak MB, Korcha RA, Greenfield TK. Child physical and sexual abuse: a comprehensive look at alcohol consumption patterns, consequences, and dependence from the National Alcohol Survey. Alcohol Clin Exp Res 2011;35:317–25.

[9] Casswell S, You RQ, Huckle T. Alcohol's harm to others: reduced wellbeing and health status for those with heavy drinkers in their lives. Addiction 2011;106:1087–94.

[10] Laslett AM, Room R, Ferris J, Wilkinson C, Livingston M, Mugavin J. Surveying the range and magnitude of alcohol's harm to others in Australia. Addiction 2011;106:1603–11.

[11] Livingston M, Wilkinson C, Laslett AM. Impact of heavy drinkers on others' health and well-being. J Stud Alcohol Drugs 2010;71:778–85.

[12] Rehm J, Rossow I. The impact of alcohol consumption on work and education. In: Klingemann H, Gmel G, eds. Mapping the social consequences of alcohol consumption. Dordrecht: Kluwer Academic Publishers, 2001:67–77.

[13] Esser MB, Gururaj G, Rao GN *et al*. Harms from alcohol consumption by strangers in five Indian states and policy implications. Drug Alcohol Rev 2017;36:682–90.

[14] Callinan S. Alcohol's harm to others: Quantifying a little or a lot of harm. Int J Alcohol Drug Res 2014;3:127–33.

[15] Room R, Ferris J, Laslett AM, Livingston M, Mugavin J, Wilkinson C. The drinker's effect on the social environment: a conceptual framework for studying alcohol's harm to others. Int J Environ Res Public Health 2010;7:1855–71.

[16] Room R. Concepts and items in measuring social harm from drinking. J Subst Abuse 2000;12:93–111.

[17] Laslett AM, Ferris J, Dietze P, Room R. Social demography of alcohol-related harm to children in Australia. Addiction 2012;107:1082–9.

[18] Fillmore KM. The social victims of drinking. Br J Addict 1985;80:307–14.

[19] Room R, Janca A, Bennett LA, Schmidt L, Sartorius N. WHO cross-cultural applicability research on diagnosis and assessment of substance use disorders: an overview of methods and selected results. Addiction 1996;91:199–220.

[20] Schmidt L, Room R. Cross-cultural applicability in international classifications and research on alcohol dependence. J Stud Alcohol 1999;60:448–62.

[21] Room R. Taking account of cultural and societal influences on substance use diagnoses and criteria. Addiction 2006;101:31–9.

[22] Wilsnack SC, Greenfield TK, Bloomfield K. The GENAHTO Project (Gender and Alcohol's Harm to Others): design and methods for a multinational study of alcohol's harm to persons other than the drinker. Int J Alcohol Drug Res 2018;7:37–47.

[23] Alcohol Research Group/Public Health Institute, USA GENAHTO – a multinational collaboration. Alcohol's harm to others: multinational cultural contexts and policy implications. 2017. Available at: http://genahto.org/ (accessed 15 April 2021).

[24] Stochl J, Jones PB, Croudace TJ. Mokken scale analysis of mental health and well-being questionnaire item responses: a non-parametric IRT

method in empirical research for applied health researchers. BMC Med Res Methodol 2012;12:74.
[25] Fischer GH, Molenaar IW. Rasch models: foundations, recent developments, and applications. New York: Springer Science & Business Media, 2012.
[26] Northwestern University Health measures - transforming how health is measured. Northwestern University, 2021. Available at: https://www.healthmeasures.net/index.php (accessed 15 April, 2021).
[27] Rizopoulos D. ltm: an R package for latent variable modeling and item response theory analyses. J Stat Softw 2006;17:1–25.
[28] Choi SW, Gibbons LE, Crane PK. Lordif: an R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. J Stat Softw 2011;39:1–30.
[29] Rossow I. How well do survey studies capture alcohol's harm to others? Subst Abuse 2015;9:99–106.
[30] Carvalho S, White H. Combining the quantitative and qualitative approaches to poverty measurement and analysis: the practice and the potential. Washington D.C.: The World Bank, 1997.
[31] World Bank Current classification by income. Available at: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups (accessed 15 April, 2021).
[32] World Bank Adjusted net national income per capita (current US $). Available at: https://data.worldbank.org/indicator/NY.ADJ.NNTY.PC.CD (accessed 15 April, 2021).
[33] Finder (US Company) The cost of a pint of beer around the world. Available at: https://www.finder.com/international-beer-price-map (accessed 15 April, 2021).
[34] World Health Organization Global status report on alcohol. 2004. Geneva: World Health Organization.
[35] Laslett AM, Jiang H, Kuntsche S *et al.* Cross-sectional surveys of financial harm associated with others' drinking in 15 countries: unequal effects on women? Drug Alcohol Depend 2020;211:107949.
[36] Bond JC, Roberts SCM, Greenfield TK, Korcha R, Ye Y, Nayak MB. Gender differences in public and private drinking contexts: A multi-level GENACIS analysis. Int J Environ Res Public Health 2010;7:2136-60.
[37] Casswell S, Zhang JF, Wyllie A. The importance of amount and location of drinking for the experience of alcohol-related problems. Addiction 1993;88:1527–34.
[38] Stockwell T, Lang E, Rydon P. High risk drinking settings: the association of serving and promotional practices with harmful drinking. Addiction 1993;88:1519–26.
[39] Bahler CD, Sundaram CP. Training in robotic surgery: simulators, surgery, and credentialing. Urol Clin North Am 2014;41:581–9.
[40] Marmet S, Gmel G. Alcohol's harm to others in Switzerland in the year 2011/2012. J Subst Use 2017;22:403–11.
[41] Karriker-Jaffe KJ, Room R, Giesbrecht N, Greenfield TK. Alcohol's harm to others: opportunities and challenges in a public health framework. J Stud Alcohol Drugs 2018;79:239–43.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Table S1. Items with DIF, number of country pairings with DIF, pairings with DIF.
Table S2. Confirmatory factor analysis by country (factor loadings, based on tetrachoric correlations), explained variance, Cronbach's α, Mokken's H.
Table S3. Severity and discrimination parameters (95% CI) by item and country.