

Determining the Sex-Specific Distributions of Average Daily Alcohol Consumption Using a Machine Learning Method: Is There a Separate Distribution for People With Alcohol Dependence?

Huan Jiang (✉ hedy.jiang@utoronto.ca)

University of Toronto <https://orcid.org/0000-0003-0898-2339>

Shannon Lange

Centre for Addiction and Mental Health

Alexander Tran

Centre for Addiction and Mental Health

Sameer Imtiaz

Centre for Addiction and Mental Health

Jürgen Rehm

Centre for Addiction and Mental Health

Research

Keywords: Alcohol consumption, machine learning, survey, Gaussian Mixture Models, clustering, alcohol use disorders, treatment utilization

Posted Date: January 12th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-142161/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

It remains unclear whether alcohol use disorders (AUDs) can be characterized by specific levels of average daily alcohol consumption. The aim of the current study was to model the distributions of average daily alcohol consumption among those who consume alcohol and those with alcohol dependence, the most severe AUD, using various machine learning techniques.

Methods

Data from Wave 1 and Wave 2 of the National Epidemiologic Survey on Alcohol and Related Conditions were used in the current analyses. Clustering algorithms were applied in order to group a set of data points that represent the average daily amount of alcohol consumed. Gaussian Mixture Models (GMMs) were then used to estimate the likelihood of a data point belonging to one of the mixture distributions. Individuals were assigned to the clusters which had the highest posterior probabilities from the GMMs, and their treatment utilization rate was examined for each of the clusters.

Results

Modeling alcohol consumption via machine learning technique was feasible. The clusters identified did not point to alcohol dependence as a separate cluster characterized by a higher level of alcohol consumption. Among both females and males with alcohol dependence, daily alcohol consumption was relatively low.

Conclusions

Overall, we found little evidence for clusters of people with the same drinking distribution, which could be characterized as clinically relevant for people with alcohol use disorders as currently defined.

Introduction

Alcohol use is a major risk factor for burden of disease (1) and alcohol use disorders (AUDs) comprise a substantial part of the alcohol-attributable disease burden globally (2). While it is obvious that an individual who completely abstains from consuming alcohol would not be diagnosed with an AUD – i.e., such use is, by definition, a necessary and sufficient cause (see (3, 4) for further discussion), there is a great deal of debate in the literature regarding whether or not these conditions can be characterized by specific levels of daily alcohol consumption. In both the Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM-5) and the International Classification of Diseases, 11th revision (ICD-11) (5, 6), AUDs are defined without any reference to the level of drinking, but rather, are based on a non-specific set of behavioral, social, psychological and physiological criteria (7, 8) – e.g., continuing to drink even though it is causing trouble with friends/family, and giving up or cutting back on activities that were once important or interesting in order to drink. While these definitions have been criticized (9, 10), level of

drinking, as an alternative key criterion, may pose other concerns (see (10, 11) for further discussion), such as:

- What would be the thresholds for minimal level of alcohol use, and for what duration would one need to consume the respective level of alcohol in order for a diagnosis to be justified?
- Would heavy use over time be sufficient for a diagnosis on its own, or would additional criteria, such as withdrawal or tolerance, be necessary, and should brain function also be considered (8)?

As for thresholds, a similar approach could be used as the one used for hypertension (12). Even though blood pressure is a continuous variable, and the higher the blood pressure, the higher the chance for various disease categories (13), expert committees do not seem to have a problem agreeing on the blood pressure levels for which an intervention is required (in other words, to determine the threshold, when the blood pressure level is high enough to be considered a disease; e.g., see (14)). However, given the modern techniques available for analyzing distributions, more empirical-based methods may differentiate seemingly continuous distributions with only one maximum (15, 16). When it comes to alcohol use, these techniques assume that the supposed unimodal distribution is a mixture of different groups of individuals with different distributions of alcohol use, which can be detected via machine learning techniques. For instance, clustering, an unsupervised learning technique, is often used to find clusters of points that appear close together (17, 18).

In the current study, we have taken this approach and have hypothesized that: 1) the (sex-specific) distributions of average daily alcohol consumption among those who consume alcohol can be described as mixture models and thus, are best represented by more than one distribution; 2) individuals with alcohol dependence, the most severe AUD, can be characterized by one of these distributions of average daily alcohol consumption; and 3) that treatment utilization will be associated with distributions of daily alcohol consumption among individuals with alcohol dependence. All analyses were sex-specific given the different body composition and the neurobiological processes of alcohol use and AUDs of males and females (19), as well as the sex-differences in the quantity and frequency of alcohol use (20) and treatment utilization rates among individuals with an AUD (21).

Methods

Data Source

The current analysis is based on data from Wave 1 and Wave 2 of the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC), designed and sponsored by the National Institute on Alcohol Abuse and Alcoholism, conducted in 2001-02 and 2004-05, respectively. The NESARC sample represents the civilian, noninstitutionalized adult population of the United States(22). The surveys were conducted using face-to-face, computer-assisted, and in-home interviews. One randomly selected adult (aged 18 years or older) from each sampled household was invited to participate. The overall response rate was

81.0% for Wave 1, for a total sample size of 43,093. Among those, 34,653 (80.4%) were followed-up in Wave 2 (8,840 participants were lost to follow-up).

The NESARC samples were weighted to adjust for probabilities of selection biases and nonresponse. Calibration was applied to match the target population based on the 2000 census. Details regarding to the sampling, weighting and calibration have been described elsewhere (23, 24).

Measures

Daily alcohol consumption

The NESARC contains detailed questions about the drink types, frequency of drinking, quantity and size of drinks consumed during the past 12 months. The amount of pure alcohol in each drink was calculated using an ethanol conversion factor, which accounts for the proportion of pure alcohol in the different types of drinks (23, 24). The average daily volume of pure alcohol consumption in grams during the past 12 months (referred to as daily alcohol consumption herein) was then calculated by dividing the total alcohol consumption across all drink types by 365.

Alcohol dependence

Alcohol dependence in the past 12-months was assessed using the Alcohol Use Disorders and Associated Disabilities Interview Schedule-IV (AUDADIS-IV), based on the criteria of the fourth edition of the DSM (DSM-IV) (25, 26) .

Treatment utilization

The NESARC defines broadly alcohol treatment utilization as “seeking help for alcohol-related problems” from at least one of the following: Alcoholics/Narcotics/Cocaine Anonymous, or 12-step meeting; family services or other social service agency; alcohol/drug detoxification ward/clinic; inpatient ward of psychiatric/general hospital or community mental; outpatient clinic, including outreach and day/partial patient programs; alcohol/drug rehabilitation program; emergency room because of drinking; halfway house/therapeutic community; crisis center because of drinking; employee assistance program; clergyman, priest, or rabbi; private physician, psychiatrist, psychologist, social worker, or any other professional; and any other agency or professional. Accordingly, for Wave 1, treatment utilization was considered as the endorsement of any of the above within the past 12-months. For Wave 2, alcohol treatment utilization was ascertained using the following question: “Have you gone anywhere or seen anyone to get help because of drinking since last interview?”

Statistical Analysis

As an exploratory analysis, following traditional fitting of distributions for alcohol use (27, 28), we evaluated the fit of the Log-Normal, Gamma, and Weibull distributions to determine if the distribution of

daily alcohol consumption could be appropriately described as unimodal, using the Wave 1 survey. The three fittings were examined using the Kolmogorov-Smirnov test and the null hypothesis was rejected for all three, which suggested the possibility of a multi-modal distribution. Given the skewness of the data, a log-transformation was applied to the daily alcohol consumption variable, and the distribution was modelled and fitted using the following steps:

1. Density plots

Density plots of daily alcohol consumption were produced and the resulting graphs were used to visually identify the possible number of modes.

2. Clustering

Clustering algorithms were used to group a set of data points into clusters, so that data points in the same cluster were more similar to each other than data points in other clusters. The desired number of clusters was decided using the NBClust package, which simultaneously varies the number of clusters, the clustering method and the indices to find the optimal number of clusters for the data points (29). When the indices fail to suggest the best clustering scheme, K-means was used to select the number of desired clusters (30).

3. Gaussian Mixture Model

Given the number of clusters, Gaussian Mixture Models (GMMs) (18) were used to estimate the likelihood that a given point belonged to one of a mixture of Gaussian distributions. The mixture distribution can be represented by writing the distribution function (F) as a sum:

$$F(x) = \sum_{i=1}^k w_i P_i(x)$$

Where k is the number of clusters, and x represents the data points and weights $\sum_{i=1}^k w_i = 1$.

$P(x)$ were assumed to follow Gaussian distributions. For each distribution, there are two parameters to describe the shape of the clusters: the mean and the standard deviation. The parameters were estimated via the Expectation-Maximization algorithm. There are two key advantages to using GMMs. Firstly, GMMs are more flexible in terms of cluster covariance. Secondly, since GMMs use probabilities, each data point can have multiple clusters. Therefore, if a data point is located in the middle of two overlapping distributions, its class can be defined by a mixed membership. The Bayesian Information Criterion was used to assess model fit. Sex-specific models were fit and visualized, as well as separate models for those with alcohol dependence.

Lastly, Wave 2 data were used to test if the respective parameter distributions could be described using GMMs were consistent with the distributions in Wave 1. In addition, an analysis using the same

statistical approach, as described above, was performed on Wave 1 and Wave 2 data combined to investigate the distributions among those individuals with alcohol dependence in both waves.

Individuals were assigned to the cluster for which they had the highest posterior probability from the GMMs. Treatment utilization rate were then calculated for each of the clusters based on lifetime treatment (seeking treatment prior to Wave 2) or any recent treatment (seeking treatment between 12 months prior to Wave 1 and Wave 2).

Results

There were more females than males in both Wave 1 and Wave 2 (52.1% vs. 47.9%; Table 1). About 4% of participants were diagnosed with alcohol dependence in both Waves (3.8% in Wave 1, and 4.4% in Wave 2). Among them, approximately two out of three were male and one out of three were female, in both waves. The density plots suggested that the daily alcohol consumption data might be coming from two or more different clusters, which were different for males and females, and for those individuals with and without alcohol dependence (Figure 1). As expected, male drinkers, as well as males with alcohol dependence consumed higher amounts of alcohol daily, on average, compared to female drinkers and females with alcohol dependence, respectively.

The log-transformed daily alcohol consumption of all individuals in Wave 1 was best described by a three-component GMM for both females and males (Figure 2). The means of the three Gaussian clusters were 0.03, 0.17 and 3.34 grams per day for females, and 0.03, 0.35 and 7.48 grams per day for males. In contrast, a two-component GMM best described the log-transformed daily alcohol consumption for both females and males with alcohol dependence. The means of the two Gaussian clusters were 0.28 and 12.82 grams per day for females with alcohol dependence, with only 2.2% of females being included in the first cluster and 97.8% included in the second cluster, and 6.12 and 35.80 grams per day for males with alcohol dependence, with 16.1% of males being included in the first cluster and 83.9% included in the second cluster.

Wave 2 yielded similar results to Wave 1 in that the log-transformed daily alcohol consumption of both females and males was best described by three-component GMMs (with means of 0.03, 0.24 and 4.86 grams per day, and 0.03, 0.34 and 7.48 grams per day, respectively, for the three Gaussian distributions), and daily alcohol consumption of males with alcohol dependence being best described by a two-component GMM (with means of 6.12 and 35.80 grams per day; see Table 2 for details and Table A1 for corresponding results in drinks per day). However, in Wave 2, daily alcohol consumption of females with alcohol dependence was best described by a three-component GMM, with the means corresponding to 3.36, 10.47 and 32.07 grams per day.

Overall, the clusters identified did not point to alcohol dependence as a separate cluster characterized by a higher level of alcohol consumption. Among both females and males with alcohol dependence, daily

alcohol consumption was relatively low, with the highest mean of any cluster being around 32 and 36 grams per day, respectively – i.e., less than three US standard drinks.

Lastly, data from Wave 1 and Wave 2 were combined to investigate the sex-specific distributions of daily alcohol consumption for those with alcohol dependence in both waves. As shown in Figure 3, average daily alcohol consumption of females with alcohol dependence was best described by a two-component GMM, while average daily alcohol consumption of males with alcohol dependence was best described by a three-component GMM. For females with alcohol dependence in both waves, 3.3% of them belonged to a cluster with a mean of 0.42 grams per day, while 96.7% of them belonged to a cluster with mean of 16.97 grams per day; de facto indicating the presence of a cluster comprised of the overwhelming majority of females with alcohol dependence with a mean drinking level of less than two drinks, which is not even considered heavy drinking (e.g., (31))

For males with alcohol dependence in both waves, the means of the three Gaussian clusters were 0.74, 17.01 and 69.99 grams per day, with 1.5%, 39.9%, and 58.6% of males belonging to the respective clusters. It is only in this analyses that we identified one cluster that had a similar level of daily alcohol consumption, on average, as “typical” treatment populations (i.e., a cluster of males with alcohol dependence in both waves consumed a mean of about 70 grams of alcohol on average per day). Table 3 shows the characteristics of individuals in each of the clusters, including the percentage who had utilized treatment, their age at Wave 1 and daily alcohol consumption in grams. Around 36.0% of males in the third cluster utilized alcohol-related treatment, compared with 21.5% of males in the second cluster. In other words, we identified a cluster of males who could be described as those with alcohol dependence requiring treatment. Whereas, the first, very small cluster among the group of males with alcohol dependence in both waves, likely consists of people who recently completed treatment, and were still abstinent. Table A2 shows the further details of those three clusters.

Discussion

A procedure for deriving different distributions of daily alcohol consumption based on a statistical clustering methodology was presented and explored. This procedure allowed for the quantitative comparison of the distributions between surveys conducted on the same individuals at different time points. Overall, we found little evidence for clusters of people with the same drinking distribution, which could be characterized as clinically relevant for people with AUDs, as currently defined. Before we discuss the results further, we would like to discuss a few potential limitations of the approach taken.

To begin, the selection of number of clusters was done through calculating multiple indices and the clustering scheme with the most agreement was adopted. When the sample size was small, the indices were less likely to agree with each other. In the case of algorithm failure and limiting computing environment, the K-means method was used to choose an appropriate number of clusters. This method assumes that all clusters are equally sized and have the same variance. When analyzing average daily

alcohol consumption for females and males in the overall sample, the sample sizes were large and the variances of all clusters were well balanced. However, the clusters for females and males with alcohol dependence in both Wave 1 and Wave 2, given the relatively small sample sizes, may not have been stable enough to be generalizable.

GMMs are primarily used in modeling populations with multiple distributions and have gained prominence within the model-based clustering framework. Using GMMs, we are able to identify occasional drinkers, light drinkers and heavy drinkers. Given the skewed distribution of daily alcohol consumption, log transformation was used to make data conform to normality, which was a requirement for the methodology used. However, it may be more difficult to interpret the findings that are based on transformed data with respect to the hypotheses of interest.

Lastly, the study was based on a survey, where the assessment of alcohol consumption was based on self-reports, which are known to underestimate true consumption levels (32, 33), either due to restricted sampling frame or due to reporting biases (34). It is unclear, how these biases impacted the current findings, but it is very likely, that heavy alcohol consumers were not part of the sample (35), which is exemplified by the fact that the homeless, institutionalized, or members of the army living on base were not included. Finally, as with all surveys, the NESARC had high, but less than perfect participation rates, and some loss to follow-up, which could impact the results.

Of all the clusters identified, only the cluster with the highest average daily alcohol consumption among males with alcohol dependence in both survey waves, showed drinking levels likely seen in treatment populations in North America (e.g., see (36, 37)), with European samples often showing higher levels (38, 39). In summary, we did not identify a cluster which could be characterised as AUD among the general drinking population or individuals with alcohol dependence at one time point only, but did identify a cluster among males with alcohol dependence in both waves.

There are a few reasons for this result, which is different from other diseases and underlying biomarkers – for example, blood glucose levels and diabetes (e.g., see (40)). First, the current criteria for AUDs are very inclusive (7), and likely do not represent need for treatment intervention, and may not grasp what has been traditionally understood as “addiction”, often defined by treatment populations (4). Consider the following: in a treatment-based sample, lifetime alcohol dependence was indeed stable, with approximately 90.5% for females and 94.7% for males, whereas in a community-based sample the stability of lifetime alcohol dependence was only 27.5% and 64.7% for females and males, respectively (41). The most important characteristic that determined diagnostic stability was severity. Thus, a diagnosis which by definition should not change, was stable in severe, treatment-seeking cases, but not in general population cases of alcohol dependence, and alcohol dependence is already the more severe AUD in the DSM-IV. In other words, measurement of AUD in the general population picks up a lot of very mild cases that are not necessary indicative of alcohol addiction or problem-drinking. Cases in which individuals are consuming alcohol more regularly appear relatively mild given that they are often forgotten by respondents when asked the same questions later.

However, the question “why can we not pick up stable groups of heavy or very heavy drinkers (42) with the current methodology” remains. We can offer three potential reasons here: first, that many people with severe alcohol dependence fall out of the sampling frame of general population surveys, as they are not living in households sampled, but are homeless or institutionalized (35). Second, they may not participate even if they had been part of the sampling frame (43). Or, third, people with an AUD do not distinctively differ in their level of drinking as a subgroup of the general population, which would be in line of theories such as the one brought forward by Ledermann (15, 16, 44). It would be important to clarify these questions, even though it is not easy given the current status of general population surveys (34).

Conclusions

The machine learning procedure proved feasible for modeling average daily alcohol consumption based on survey data and allowed for the quantitative differentiation of the distributions among the study populations. We concluded that AUDs as currently defined could not be described by a group of people based on average level of drinking. This conclusion could reflect a problem with the underlying definitions which may be too unspecific and wide (4). However, it may also reflect the fact, that people with AUD do not represent a distinct category with respect to alcohol use, but one group within a continuum of use, as described by Skog(16) or Lederman (15). The answer to this question not only would increase current knowledge, but also has practical implications for conceptualizing disease and clinical care.

Abbreviations

alcohol use disorders (AUDs)

the Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM-5)

the International Classification of Diseases, 11th revision (ICD-11)

the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC)

the Alcohol Use Disorders and Associated Disabilities Interview Schedule-IV (AUDADIS-IV)

the fourth edition of the DSM (DSM-IV)

Gaussian Mixture Models (GMMs)

Declarations

-Ethics approval and consent to participate

Not applicable

-Consent to publication

The manuscript contains no individual person's data in any form.

- Availability of data and material

The R Code used to analyze and compute variables for the current study are available from the corresponding author upon reasonable request. The data may be obtained by request through the National Institute on Alcohol Abuse and Alcoholism.

- Competing interests

No competing interests.

- Funding

Not applicable

- Authors' contributions

HJ contributed to the conception and design of the study, prepared the dataset, conducted the statistical analyses and led the writing of the manuscript. SL contributed to the design and conception of the study. AT contributed to the design and assisted with the statistical analysis. SI contributed to the design and provided the original dataset. JR led the conception and design of the study. All authors contributed to the drafting of the manuscript and introduced revisions, approved the final version of the text, and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolve

- Acknowledgements

Not applicable

References

1. Shield K., Manthey J., Rylett M., Probst C., Wettlaufer A., Parry C. D. et al. National, regional, and global burdens of disease from 2000 to 2016 attributable to alcohol use: a comparative risk assessment study, *The Lancet Public Health* 2020: 5: e51-e61.
2. Rehm J., Shield K., Gmel G., Rehm M., Frick U. Modeling the impact of alcohol dependence on mortality burden and the effect of available treatment interventions in the European Union, *European Neuropsychopharmacology* 2013: 23: 89-97.
3. Rehm J., Gmel G. E., Sr., Gmel G., Hasan O. S. M., Imtiaz S., Popova S. et al. The relationship between different dimensions of alcohol use and the burden of disease-an update, *Addiction* 2017: 112: 968-1001.

4. Markus Heilig J. M., Diana Martinez, Jürgen Rehm, Lorenzo Leggio, Louk J.M.J. Vanderschuren. Addiction as a brain disease revised: why it still matters, and the need for consilience, Neuropsychopharmacology 2020.
5. World Health Organization. ICD-11 for Mortality and Morbidity Statistics 2018.
6. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (5th edition) Philadelphia, USA: American Psychiatric Association; 2013.
7. Carvalho A. F., Heilig M., Perez A., Probst C., Rehm J. Alcohol use disorders, Lancet 2019: 394: 781-792.
8. Rehm J., Heilig M., Gual A. ICD-11 for Alcohol Use Disorders: Not a Convincing Answer to the Challenges, Alcohol Clin Exp Res 2019: 43: 2296-2300.
9. Martin C. S., Langenbucher J. W., Chung T., Sher K. J. Truth or consequences in the diagnosis of substance use disorders, Addiction 2014: 109: 1773-1778.
10. Rehm J., Marmet S., Anderson P., Gual A., Kraus L., Nutt D. J. et al. Defining substance use disorders: do we really need more than heavy use?, Alcohol and alcoholism 2013: 48: 633-640.
11. Rehm J., Anderson P., Gual A., Kraus L., Marmet S., Nutt D. et al. The tangible common denominator of substance use disorders: a reply to commentaries to Rehm et al.(2013), Alcohol and alcoholism 2014: 49: 118-122.
12. Nutt D. J., Rehm J. Doing it by numbers: a simple approach to reducing the harms of alcohol, Journal of Psychopharmacology 2014: 28: 3-7.
13. Singh G. M., Danaei G., Farzadfar F., Stevens G. A., Woodward M., Wormser D. et al. The age-specific quantitative effects of metabolic risk factors on cardiovascular diseases and diabetes: a pooled analysis, PloS one 2013: 8: e65174.
14. Unger T., Borghi C., Charchar F., Khan N. A., Poulter N. R., Prabhakaran D. et al. 2020 International Society of Hypertension global hypertension practice guidelines, Hypertension 2020: 75: 1334-1357.
15. Ledermann S. Alcool, Alcoolisme, Alcoolisation , Vol. I Paris, France: Presses Universitaires de France; 1956.
16. Skog O. J. The collectivity of drinking cultures: A theory of the distribution of alcohol consumption, British Journal of Addiction 1985: 80: 83-99.
17. Yu G., Sapiro G., Mallat S. Solving inverse problems with piecewise linear estimators: From Gaussian mixture models to structured sparsity, IEEE Transactions on Image Processing 2011: 21: 2481-2499.
18. Bishop C. M. Pattern recognition and machine learning: springer; 2006.
19. Becker J. B., Koob G. F. Sex differences in animal models: focus on addiction, Pharmacological reviews 2016: 68: 242-263.
20. Grant B. F., Chou S. P., Saha T. D., Pickering R. P., Kerridge B. T., Ruan W. J. et al. Prevalence of 12-month alcohol use, high-risk drinking, and DSM-IV alcohol use disorder in the United States, 2001-2002 to 2012-2013: results from the National Epidemiologic Survey on Alcohol and Related Conditions, JAMA psychiatry 2017: 74: 911-923.

21. Ilgen M. A., Price A. M., Burnett-Zeigler I., Perron B., Islam K., Bohnert A. S. et al. Longitudinal predictors of addictions treatment utilization in treatment-naïve adults with alcohol use disorders, *Drug and Alcohol Dependence* 2011; 113: 215-221.
22. Imtiaz S., Loheswaran G., Le Foll B., Rehm J. Longitudinal alcohol consumption patterns and health-related quality of life: Results from the National Epidemiologic Survey on Alcohol and Related Conditions, *Drug and alcohol review* 2018; 37: 48-55.
23. Grant B. F., Dawson D. A., Stinson F. S., Chou P. S., Kay W., Pickering R. 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* 2003; 71: 7-16.
24. Huang B., Grant B. F., Dawson D. A., Stinson F. S., Chou S. P., Saha T. D. et al. Race-ethnicity and the prevalence and co-occurrence of Diagnostic and Statistical Manual of Mental Disorders, alcohol and drug use disorders and Axis I and II disorders: United States, 2001 to 2002, *Comprehensive psychiatry* 2006; 47: 252-257.
25. Edition F. Diagnostic and statistical manual of mental disorders, Am Psychiatric Assoc 2013.
26. Hasin D. S., Stinson F. S., Ogburn E., Grant B. F. 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* 2007; 64: 830-842.
27. Kehoe T., Gmel G., Shield K. D., Gmel G., Rehm J. Determining the best population-level alcohol consumption model and its impact on estimates of alcohol-attributable harms, *Population health metrics* 2012; 10: 6.
28. Rehm J., Kehoe T., Gmel G., Stinson F., Grant B., Gmel G. Statistical modeling of volume of alcohol exposure for epidemiological studies of population health: the US example, *Population Health Metrics* 2010; 8: 3.
29. Charrad M., Ghazzali N., Boiteau V., Niknafs A. Determining the best number of clusters in a data set, *J Stat Softw* 2014.
30. Hartigan J. A., Wong M. A. Algorithm AS 136: A k-means clustering algorithm, *Journal of the royal statistical society series c (applied statistics)* 1979; 28: 100-108.
31. Agency E. M. Guideline on the development of medicinal products for the treatment of alcohol dependence, EMEA/CHMP/EWP/20097/2008 2010.
32. Midanik L. T. Validity of self-reported alcohol use: A literature review and assessment, *British journal of addiction* 1988; 83: 1019-1029.
33. Rehm J., Klotsche J., Patra J. Comparative quantification of alcohol exposure as risk factor for global burden of disease, *International Journal of Methods in Psychiatric Research* 2007; 16: 66-76.
34. Rehm J., Kilian C., Rovira P., Shield K. D., Manthey J. The elusiveness of representativeness in general population surveys for alcohol, *Drug and alcohol review* 2020.
35. Shield K. D., Rehm J. Difficulties with telephone-based surveys on alcohol consumption in high-income countries: the Canadian example, *International journal of methods in psychiatric research*

2012: 21: 17-28.

36. Roesner S., Hackl-Herrwerth A., Leucht S., Leherth P., Vecchi S., Soyka M. Acamprosate for alcohol dependence, Cochrane Database of Systematic Reviews 2010.
37. Roesner S., Hackl-Herrwerth A., Leucht S., Vecchi S., Srisurapanont M., Soyka M. Opioid antagonists for alcohol dependence, Cochrane database of systematic reviews 2010.
38. Rehm J., Manthey J., Struzzo P., Gual A., Wojnar M. Who receives treatment for alcohol use disorders in the European Union? A cross-sectional representative study in primary and specialized health care, European Psychiatry 2015: 30: 885-893.
39. Rehm J., Allamani A., Aubin H.-J., Della Vedova R., Elekes Z., Frick U. et al. People with alcohol use disorders in specialized care in eight different European countries, Alcohol and Alcoholism 2015: 50: 310-318.
40. Fan J., May S. J., Zhou Y., Barrett-Connor E. Bimodality of 2-h plasma glucose distributions in whites: the Rancho Bernardo study, Diabetes care 2005: 28: 1451-1456.
41. Culverhouse R., Bucholz K. K., Crowe R. R., Hesselbrock V., Nurnberger J. I., Porjesz B. et al. Long-term stability of alcohol and other substance dependence diagnoses and habitual smoking: an evaluation after 5 years, Archives of general psychiatry 2005: 62: 753-760.
42. Rehm J., Guiraud J., Poulmais R., Shield K. D. Alcohol dependence and very high risk level of alcohol consumption: a life-threatening and debilitating disease, Addiction biology 2018: 23: 961-968.
43. Meiklejohn J., Connor J., Kypri K. The effect of low survey response rates on estimates of alcohol consumption in a general population survey, PloS one 2012: 7: e35527.
44. Rose G. The population strategy of prevention, The strategy of preventive medicine 1992: 95-106.

Tables

Table 1. Weighted demographic information for Wave 1 and Wave 2

	Wave 1	Wave 2
	N (%) ^a	N (%)
Gender		
Females	22,443 (52.1)	18,048 (52.1)
Males	20,650 (47.9)	16,605 (47.9)
Age		
≤20	1,698 (3.9)	0 (0.0)
20-40	16,358 (38.0)	12,144 (35.1)
41-70	20,000 (46.4)	17,592 (50.8)
71+	5,038 (11.7)	4,913 (14.2)
Alcohol dependence, past 12-months	1,640 (3.8)	1,521 (4.4)
Females	521 (31.8)	478 (31.4)
Males	1,119 (68.2)	1,043 (68.6)
Alcohol treatment utilization ^b	246 (0.6)	473 (1.4)

^a N is adjusted by the sample weight.

^b Alcohol treatment utilization is broadly defined as “seeking help for alcohol-related problems” from at least one of 13 sources (e.g., health care facilities, alcohol-related programs, family or social services, employee assistance program or religious services).

Table 2. Summary of the sex-specific Gaussian Mixture Models for drinkers overall, and for individuals with alcohol dependence using Wave 2 data

	Females	Males	Females with alcohol dependence	Males with alcohol dependence
N	10,970	11,897	207	364
First component				
Average daily consumption (g)	0.03	0.03	3.36	6.12
Weight	0.06	0.03	0.60	0.16
Second component				
Average daily consumption (g)	0.24	0.34	10.47	35.80
Weight	0.52	0.31	0.29	0.84
Third component				
Average daily consumption (g)	4.86	7.48	32.07	n/a
Weight	0.42	0.66	0.12	n/a

g: grams; n/a: not applicable, as model only yielded two distributions.

Table 3. Cluster descriptions for those with alcohol dependence in both Wave 1 and Wave 2

	N	Treatment utilization, lifetime (%)	Treatment utilization, past 12-months at Wave 1 and/or between Wave 1 and Wave 2 (%)	Age (years) at Wave 1 – mean (SD)	Average daily alcohol consumption (g) – mean (SD)
Females					
Cluster 1	4	0.0	0.0	20.98 (5.84)	0.47 (0.25)
Cluster 2	126	31.4	21.5	34.94 (11.34)	25.67 (27.99)
Males					
Cluster 1	5	86.3	86.3	31.24 (3.10)	0.75 (0.16)
Cluster 2	131	35.7	21.5	31.31 (11.12)	17.09 (7.50)
Cluster 3	201	55.9	36.0	32.98(12.03)	84.43 (49.46)

g: grams; SD: standard deviation

Figures

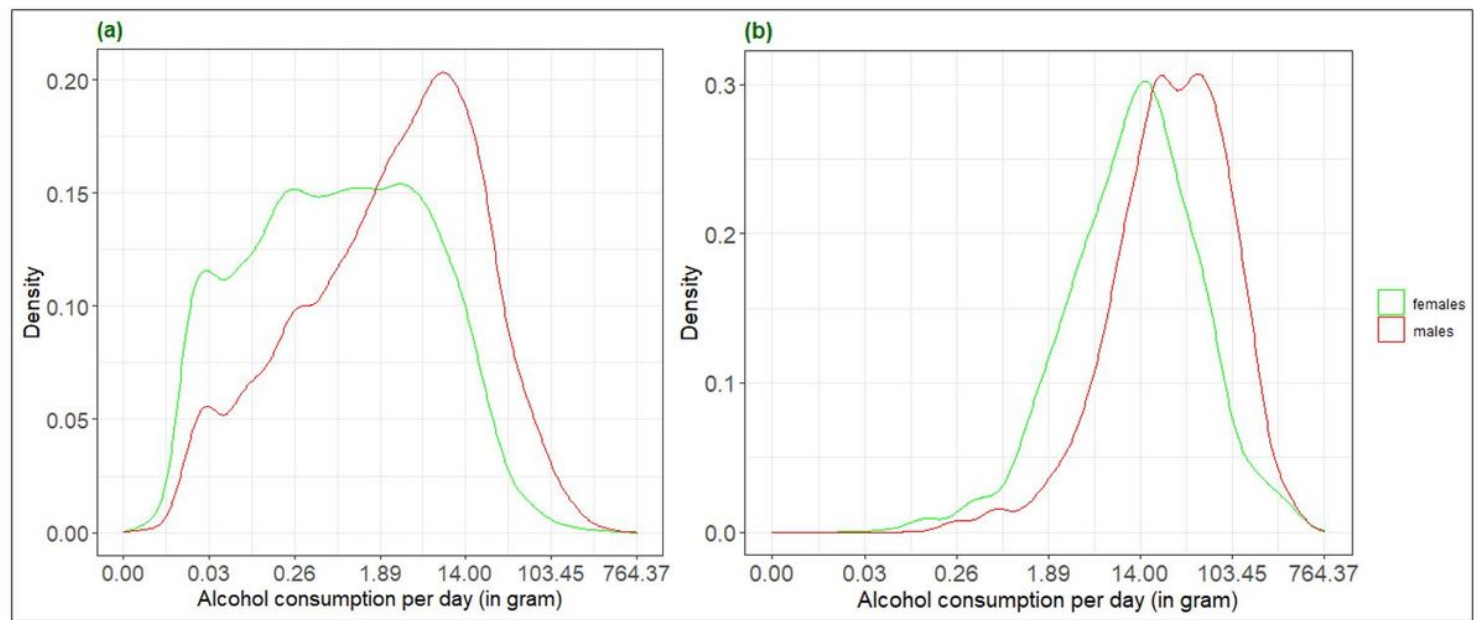


Figure 1

Density plots of daily alcohol consumption (log-transformed) from Wave 1 for (a) all individuals and (b) those individuals with alcohol dependence, by sex

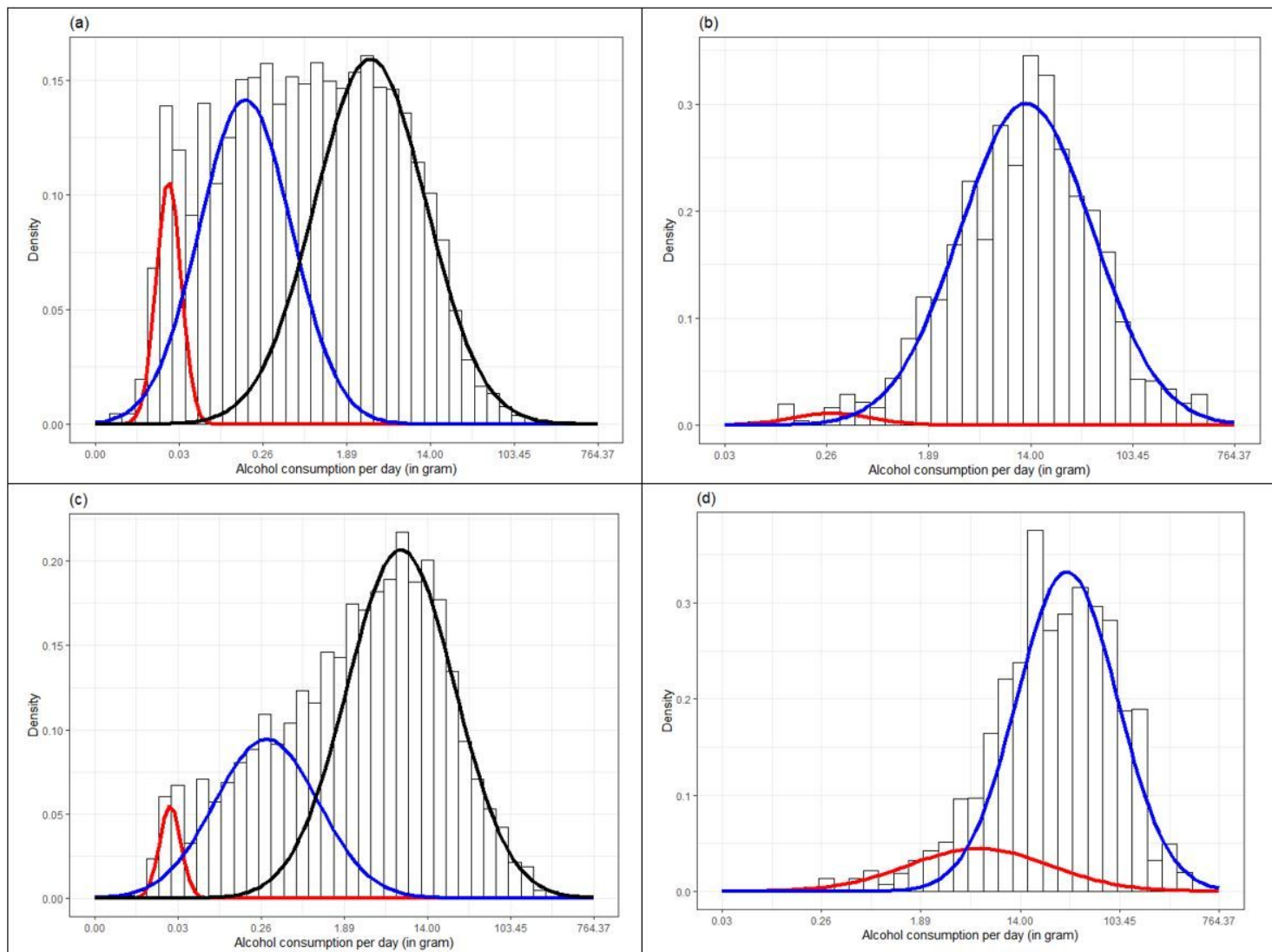


Figure 2

Gaussian Mixture Models for (a) females, (b) females with alcohol dependence, (c) males, and (d) males with alcohol dependence, using Wave 1 data.

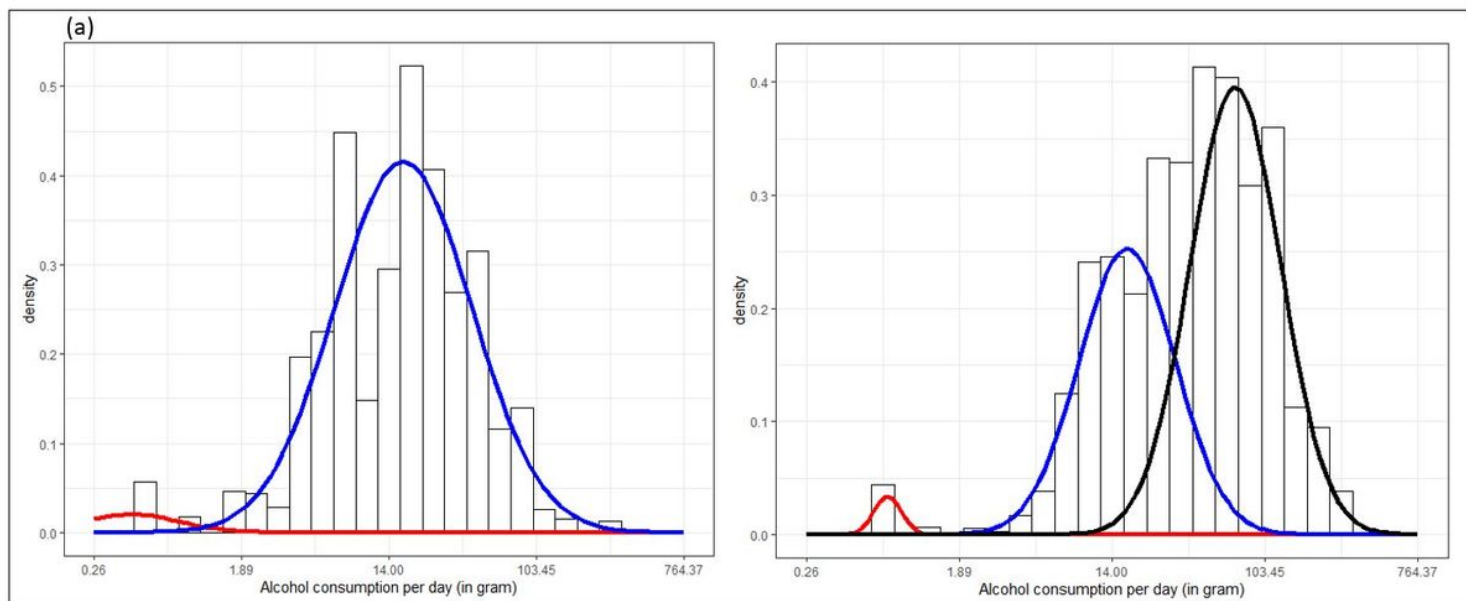


Figure 3

Gaussian Mixture Models for (a) females and (b) males with alcohol dependence in both Wave 1 and Wave 2