This is the peer reviewed version of the following article:

Barratt, M. J., & Lenton, S. (2015). Representativeness of online purposive sampling with Australian cannabis cultivators. *International Journal of Drug Policy, 26*, 323–326.

which has been published in final form at
https://doi.org/10.1016/j.drugpo.2014.10.007

This article may be used for non-commercial purposes in accordance with Elsevier's Article Sharing policy and Hosting policy

© 2015. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/

Representativeness of online purposive sampling with Australian cannabis cultivators

Monica J. Barratt[a,b], Simon Lenton[b]

[a] Drug Policy Modelling Program, National Drug and Alcohol Research Centre, UNSW Australia, Sydney NSW 2052, Australia. E-mail: m.barratt@unsw.edu.au

[b] National Drug Research Institute, Faculty of Health Sciences, Curtin University, GPO Box U1987, Perth WA 6845, Australia.

Correspondence: Dr Monica Barratt, Drug Policy Modelling Program, National Drug and Alcohol Research Centre, UNSW Australia, Sydney NSW 2052, Australia. Phone: +61407778938. E-mail: m.barratt@unsw.edu.au

Word count (body of text) [limit of 2000] = 1991

Representativeness of online purposive sampling with Australian cannabis cultivators

Keywords [3-6]: Hidden population, Hard-to-reach, Internet, Recruitment, Sampling

Abstract

Background: Digital research methods have become useful tools in the study of hidden populations. Large samples are often obtained, but their external validity can be unclear. This paper compares an online purposive sample of Australian cannabis growers with a matched sub-sample accessed from a general population survey.

Methods: We compared the demographic and drug use characteristics of recent cannabis users who report recent cannabis cultivation from a probability (National Drug Strategy Household Survey, n=100) and purposive online sample (Global Cannabis Cultivation Research Consortium, n=327) using nonparametric (bootstrap) and meta-analysis techniques.

Results: The samples were not statistically significantly different by age, indigenous status, employment, completion of any educational qualification, living alone or with partner, daily cannabis use and recent other illicit drug use. The purposive online sample was more likely to report being male, residing in a regional/remote area and completion of a university qualification, while the probability sample was more likely to report first using cannabis under the age of 16 years and living with children.

Conclusion: While our sample cannot be understood as representative, these findings provide greater confidence in our purposive sampling methodology that uses a wide range of internet and traditional recruitment sources.

Paper type: Short report

Abstract word count [limit of 200 words]: 196

## Introduction

Accessing people for research purposes who cultivate cannabis is difficult. The illegality of cannabis cultivation in most countries makes it more likely that such populations are motivated to remain hidden to avoid potential legal and social consequences as a result of their activities being revealed to others, especially law enforcement. The hidden nature of these populations poses problems for researchers who aim at collecting a representative sample of respondents, in order to be more confident that their findings are externally valid. Probability sampling methods are limited when applied to the study of hidden populations, because: (a) response rates for general population surveys are decreasing (Groves, 2006); (b) hidden populations are often more likely to be excluded from probability survey sampling frames (Zhao, Stockwell, & Macdonald, 2009); and (c) probability sampling methods are expensive, especially when low-prevalence behaviours are targeted (Kakinami & Conner, 2010). To some extent, these limitations can be addressed through the use of purposive sampling through digital technologies. Purposive sampling, which requires researchers to develop a situated knowledge of the field site and rapport with target population members, has a long history in the drugs field (e.g., Braunstein, 1993), but has traditionally been associated with small samples. However, when purposive sampling is combined with digital communication methods, researchers have successfully engaged large samples who are otherwise hard to reach at relatively low cost (Miller & Sønderlund, 2010). Such large samples produce datasets that lend themselves to quantitative analysis, yet the external validity of such analyses largely remains unknown, as it is usually neither possible to determine a response rate nor the characteristics of non-responders or potential responders who were out of scope. Despite the usual caveats published with them that their findings should not be generalised, in the absence of other information about hidden populations, these findings tend to be assigned greater external validity than they may warrant. Thus, we have recently argued that comparisons between matched sub-samples from both purposive and probability survey methodologies should, where possible, be conducted and published alongside quantitative analyses from purposive samples (Barratt, Ferris, & Lenton, 2014). Such comparisons may allow the readers and producers of research to have greater confidence to evaluate the external validity of purposive samples, and interpret the representativeness of the resultant findings.

The Global Cannabis Cultivation Research Consortium (GCCRC), as outlined in this volume, has accessed the largest international sample of cannabis growers to date through online purposive sampling methodologies. How can we assess the representativeness of this sample? Representativeness is difficult to assess because representative household surveys that access cannabis cultivators are rare. Australia, Finland and Denmark were the only countries from the GCCRC sample countries where their national representative surveys asked about growing and therefore produced subsamples of growers to which could potentially be compared with our online samples. However, the Finland (n=30, Pekka Hakkarainen, personal communication) and Denmark (n=22, Vibeke Asmussen Frank, personal

communication) samples of cannabis growers from their representative national surveys were too small to make meaningful comparisons, so we have focused only on the Australian sample (n=100) in this paper. This paper compares the online purposive sample of Australian cannabis growers with a matched sub-sample of cannabis growers accessed from Australia's representative general population survey, the National Drug Strategy Household Survey (NDSHS, Australian Institute of Health and Welfare [AIHW], 2011).

## Methods

### *Data sources*

The Australian component of the GCCRC survey was open to respondents from July 2012 to February 2013, and attracted 574 Australian residents who reported that they were 18 years or over and had grown cannabis at least once. The online survey of cannabis cultivation practices (the International Cannabis Cultivation Questionnaire [ICCQ], Decorte et al., 2012) took approximately 15 minutes to complete. As outlined more fully in Barratt el al. this volume and Barratt et al. (2012), we employed a large variety of recruitment methods to access this hidden population. The eligible Australian sample was recruited via cannabis organisations/websites/forums (22%), news articles (20%), radio (17%), Facebook (14%), other drug organisations/websites/forums (10%), and a number of other sources, including friend referral (all other 17%). Engagement with cannabis growers who participated in online groups was sought, rather than simply advertising in places where online group members gathered (see Barratt & Lenton, 2010). The study was approved by the Curtin Human Research Ethics Committee (NDRI-01-2012).

The 2010 NDSHS was the tenth survey conducted every 2-3 years since 1985 which aimed to determine the prevalence of licit and illicit drug use among Australian residents aged 14 years and over. The 2010 survey comprised 26,648 complete usable questionnaires, with a response rate of 50.6%. The sampling excluded non-private dwellings (e.g. hospitals, prisons, drug and alcohol rehabilitation centres, refuges, university residences, etc.) and the homeless (AIHW, 2011).

To make valid comparison between datasets, we matched the sub-samples as much as possible. The NDSHS general population survey asked all recent (last 12 month) cannabis users where they usually obtained cannabis (single response). Those who reported that they usually obtained cannabis from 'grow my own' (n=100) were included as growers for the current analysis. All of these respondents were over 18 years of age, which was also the lower age eligibility criterion for the purposive sample. The matching purposive sample included those who had used cannabis in the last 12 months and had grown cannabis in the last 12 months (n=327). Therefore, one limitation is that the samples are not completely matched: some respondents to the NDSHS may have grown cannabis in the last 12 months, but not as their 'usual' source.

## *Measures*

Demographic variables available for comparison included age, sex, remoteness, indigenous status, employment status, completed educational qualifications, household structure (alone, with partner, with children). Drug variables available for comparison included age of first cannabis use (less than 16 years), daily cannabis use, and recent (last 12 month) other illicit drug use. While most variables were asked identically between the two surveys, the following items were different and this may limit their comparison: remoteness, language, employment, and household structure. Remoteness: In the ICCQ, growers were asked to estimate the remoteness of the region in which they lived (major city, inner regional, outer regional, remote, very remote), whereas in the NDSHS, residential addresses from the sampling frame were categorised directly into these options. Outer regional, remote and very remote have been collapsed for analysis. Employment: In the ICCQ, growers could select more than one option from a list of multiple responses to the question 'what is your current employment status?', whereas in the NDSHS they were asked 'which of the following best described your main current employment status?'. These variables were used to derive a dichotomous indicator of whether the grower was in paid employment. We would expect a greater proportion of the GCCRC sample to have indicated paid employment given the different question structure – e.g. a full-time student in casual employment may choose 'student' using NDSHS question but would be counted as 'employed' using the ICCQ. Household structure: The ICCQ asked 'who else lives with you in the same household?', whereas the NDSHS asked 'which category best described this household?', with indicator variables for living alone, living with partner/spouse, and living with children derived from the above. All other variables were identical or comparable. Readers can view the NDSHS questionnaire here (AIHW, 2011) and the ICCQ here (Decorte, et al., 2012).

## *Analysis*

Means for continuous variables or percentages for categorical variables are presented for each sample. Stata 11.1 (StataCorp, 2009) was used to estimate confidence intervals. Design weights and weighted numbers and percentages were reported for the NDSHS data. Non-parametric bootstrap confidence intervals were estimated around comparable estimates and proportions from the purposive sample. Subsampling was repeated 250 times to generate each estimate. While bootstrap confidence intervals are better suited to estimating confidence intervals using the convenience sample as the population (Adèr, Mellenbergh, & Hand, 2008), standard linearized confidence intervals (not shown) also produced very similar results. To compare the results from the two samples directly, we undertook an analysis using techniques typically applied in meta-analysis studies (Sterne, 2009) using an alpha level of .05.

## Results

Comparing the Australian GCCRC sample with the subsample of growers from the NDSHS sample revealed that they were not significantly different on a range of important variables. Notably, age, indigenous status, involvement in paid employment, completion of any educational qualification, proportion living alone, proportion living with partner, daily cannabis use and recent other illicit drug use were not statistically significantly different between samples. However, growers accessed through the NDSHS were more likely to be female, more likely to reside in a capital city, less likely to reside in an outer regional or remote area, less likely to have completed a university qualification, more likely to be living with children, and more likely to report first using cannabis under the age of 16, compared with the GCCRC sample.

[Insert Table 1 about here]

## Discussion

We have accessed a sample of cannabis growers that are relatively comparable with those accessed through probability household survey methodology. While this finding does not mean our sample can be understood as representative, it does lend more confidence to our purposive sampling methodology that uses a wide range of internet and traditional recruitment sources.

Demographic variables that differed may potentially be explained. Routine monitoring of internet access in Australia (Australian Bureau of Statistics [ABS], 2014) indicates that although the digital divide between males and females has disappeared, people aged over 55 and those not employed are less likely to access the internet. Internet access is still strongly associated with rising levels of educational attainment and income. These differences in overall internet access may explain the finding that the online sample was more highly educated; however, it is clear that access is only one factor – desire and motivation to complete the survey and the extent of use of online cannabis groups, websites and social media would play an important role in whether the potential respondent has the opportunity to be included. Furthermore, there is also still a gender divide favouring male participants in online drug-use communities (e.g., 76% of 897 Bluelight.org survey respondents were male; Chiauzzi, DasMahapatra, Lobo, & Barratt, 2013) which is likely to have contributed to a bias towards males in the online sample.

Regarding the household survey sample being more likely to have first tried cannabis under the age of 16, the sub-sample differences may have resulted in a more cannabis-involved group, as cannabis users whose growing was a secondary source of supply were excluded. However, this interpretation is not supported by the comparison on daily cannabis use, which, although the difference did not reach statistical significance, indicated that the online sample were heavier users of cannabis than the household sample.

The online survey reached a greater proportion of growers living in outer regional or remote areas than the household survey, even though data on internet access in Australia (ABS, 2014) indicates that urban areas have greater internet access. Perhaps growers in these more remote areas may be more likely to participate in online groups and networks of growers for support, through which they were recruited to the survey? They may also be more likely to grow cannabis in the more plentiful secluded outdoor or warehouse areas compared with urban environments. This increased reach in more remote areas may indicate a benefit of using digital research methods which have been demonstrated to more easily access such populations that are often hard to reach.

In interpreting this analysis and what it might mean more broadly for our global project, we should note that the Australian GCCRC team used a broad recruitment strategy which resulted in two fifths of its respondents being recruited through mainstream media channels, whereas some other countries recruited their sample entirely through online sources (see Table 2, Barratt et al., this volume). It is unfortunate that the data are not available to be able to make similar comparisons across all participating countries, but this also highlights the uniqueness and importance of our study.

## Acknowledgements

We would like to thank the cannabis cultivators who completed our questionnaire. Our research would not be possible without your efforts. We also thank all individuals and groups who promoted the survey, including but not limited to OZStoners.com, Bluelight.org, Shaman Australis, and Nimbin Hemp Embassy. Thanks also to Pekka Hakkarainen and Vibeke Asmussen Frank for their comments on an earlier draft of this paper. Simon Lenton and Monica Barratt, through their employment at the National Drug Research Institute at Curtin University, were supported by funding from the Australian Government under the Substance Misuse Prevention and Service Improvement Grants Fund. Monica Barratt was also supported by a National Health and Medical Research Council Early Career Fellowship (APP1070140). The confidentialised unit record file for the 2010 NDSHS was provided by the Australian Institute of Health and Welfare, the Department of Health and Ageing, and the Australian Data Archive; however, they bear no responsibility for the analyses presented within this publication or the interpretation of them.

## References

Adèr, H. J., Mellenbergh, G. J., & Hand, D. J. (2008). *Advising on research methods: A consultant's companion*. Huizen, The Netherlands: Johannes van Kessel Publishing.

Australian Bureau of Statistics (ABS). (2014). *8146.0 - Household Use of Information Technology, Australia, 2012-13*. Canberra: Author.

Australian Institute of Health and Welfare (AIHW). (2011). *2010 National Drug Strategy Household Survey* (Drug statistics series no. 25. Cat. no. PHE 145). Canberra: Author.

Barratt, M. J., Bouchard, M., Decorte, T., Asmussen Frank, V., Hakkarainen, P., Lenton, S., Malm, A., Nguyen, H., & Potter, G. R. (2012). Understanding global patterns of domestic cannabis cultivation. *Drugs and Alcohol Today, 12*, 213-221.

Barratt, M. J., Ferris, J. A., & Lenton, S. (2014). Hidden populations, online purposive sampling, and external validity: Taking off the blindfold. *Field Methods*. Epub 2014 Apr 14. doi: 10.1177/1525822X14526838

Barratt, M. J., & Lenton, S. (2010). Beyond recruitment? Participatory online research with people who use drugs. *International Journal of Internet Research Ethics, 3*, 69-86.

Braunstein, M. S. (1993). Sampling a hidden population: noninstitutionalized drug users. *AIDS Education and Prevention, 5*, 131-140.

Chiauzzi, E., DasMahapatra, P., Lobo, K., & Barratt, M. J. (2013). Participatory research with an online drug forum: a survey of user characteristics, information sharing, and harm reduction views. *Substance Use and Misuse, 48*, 661-670.

Decorte, T., Barratt, M. J., Nguyen, H., Bouchard, M., Malm, A., & Lenton, S. (2012). *International Cannabis Cultivation Questionnaire (ICCQ) (Version 1.1)* (Technical Report). Belgium: Global Cannabis Cultivation Research Consortium. http://worldwideweed.nl/iccq/

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly, 70*, 646-675.

Kakinami, L., & Conner, K. R. (2010). Sampling Strategies for Addiction Research. In P. G. Miller, J. Strang & P. M. Miller (Eds.), *Addiction Research Methods* (pp. 27-42). Chichester, UK: Wiley-Blackwell.

Miller, P. G., & Sønderlund, A. L. (2010). Using the internet to research hidden populations of illicit drug users: A review. *Addiction, 105*, 1557-1567.

StataCorp. (2009). Stata Statistical Software: Release 11. College Station, TX: StataCorp LP.

Sterne, J. A. C. (2009). *Meta-analysis in Stata: An updated collection from the Stata Journal*. College Station, TX: Stata Press.

Zhao, J., Stockwell, T. I. M., & Macdonald, S. (2009). Non–response bias in alcohol and drug population surveys. *Drug and Alcohol Review, 28*, 648-657.

Table 1

Characteristics of recent cannabis growers from an online purposive sample (GCCRC) matched with the NDSHS probability sample

| Characteristics | GCCRC Online purposive (N = 327) | | NDSHS Household probability [a] (N = 100) | | Test for mean differences | |
|---|---|---|---|---|---|---|
| | n | % (95% CI Boot) | N | % (95% CI Linear) | $M_{diff}$ (95% CI Linear) | *p* |
| Age | n = 322 | | n = 100 | | | |
| mean | 37.5 | (36.1–38.9) | 40.0 | (37.5–42.5) | 2.5% (-0.4–5.3) | .092 |
| SD (range) | 12.8 | (18–71) | 11.6 | (18–65) | | |
| Sex | | | | | | |
| Male | 290 of 326 | 89.0 (85.5–92.4) | 77 of 100 | 76.7 (68.1–85.3) | 12.3% (3.0–21.6) | .009 |
| Rurality classification | | | | | | |
| Capital city | 141 of 321 | 43.9 (38.5–49.4) | 57 of 100 | 56.9 (45.8–67.9) | 12.9% (0.6–25.2) | .006 |
| Outer regional/remote | 111 of 321 | 34.6 (29.6–39.6) | 18 of 100 | 18.2 (9.5–27.0) | 16.3% (6.3–26.4) | .001 |
| Ethnicity | | | | | | |
| Indigenous status | 10 of 319 | 3.1 (1.2–5.1) | 4 of 93 | 4.9 (0.0–9.7) | 1.7% (-3.5–7.0) | .520 |
| Current employment status | | | | | | |
| In paid employment | 227 of 325 | 69.8 (64.9–74.8) | 71 of 96 | 73.8 (64.1–83.4) | 3.9% (-7.0–14.8) | .481 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Highest qualification | | | | | | |
| Any qualification [b] | 208 of 320 | 65.0 (60.1–69.9) | 71 of 100 | 71.1 (60.8–81.4) | 6.1% (-5.3–17.5) | .291 |
| University qualification | 77 of 320 | 24.1 (19.6–28.5) | 12 of 100 | 11.9 (5.6–18.2) | 12.2% (4.5–20.0) | .002 |
| Living situation | | | | | | |
| Lives alone | 44 of 325 | 13.5 (9.6–17.5) | 16 of 95 | 16.6 (9.1–24.0) | 3.0% (-5.4–11.4) | .481 |
| Lives with partner | 199 of 325 | 61.2 (55.9–66.6) | 61 of 95 | 63.8 (53.1–74.5) | 2.6% (-9.4–14.6) | .676 |
| Lives with children | 97 of 325 | 29.8 (24.5–35.2) | 47 of 95 | 49.6 (38.2–61.1) | 19.8% (7.2–32.4) | .002 |
| Selected drug use characteristics | | | | | | |
| Cannabis first use under 16 years | 110 of 327 | 33.6 (28.5–38.8) | 48 of 100 | 48.4 (37.2–59.6) | 14.8% (2.5–27.1) | .019 |
| Daily cannabis user | 144 of 327 | 44.0 (38.1–50.0) | 32 of 99 | 32.7 (22.2–43.3) | 11.3% (0.8–23.4) | .069 |
| Recent user of other illicit drugs | 122 of 327 | 37.3 (32.3–42.3) | 39 of 98 | 39.4 (28.0–50.9) | 2.1% (-10.4–14.6) | .739 |

[a] All household survey estimates are weighted. [b] Includes trade or technical certificate or diploma, undergraduate and postgraduate qualifications.