

BMJ Open Usability of existing alcohol survey data in South Africa: a qualitative analysis

Mayara Fontes Marx,¹ Leslie London,² Nadine Harker Burnhams,³ John Ataguba²

To cite: Fontes Marx M, London L, Harker Burnhams N, *et al*. Usability of existing alcohol survey data in South Africa: a qualitative analysis. *BMJ Open* 2019;**9**:e031560. doi:10.1136/bmjopen-2019-031560

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2019-031560>).

Received 09 May 2019
Revised 21 June 2019
Accepted 12 July 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Health Science Facult, University of Cape town, Cape town, South Africa

²School of Public Health and Family Medicine, University of Cape Town, Cape Town, South Africa

³South African Medical Research Council, Tygerberg, South Africa

Correspondence to

Dr Mayara Fontes Marx;
crrmay002@myuct.ac.za

ABSTRACT

Objective This paper assesses the usability of existing alcohol survey data in South Africa (SA) by documenting the type of data available, identifying what possible analyses could be done using these existing datasets in SA and exploring limitations of the datasets.

Settings A desktop review and in-depth semistructured interviews were used to identify existing alcohol surveys in SA and assess their usability.

Participants We interviewed 10 key researchers in alcohol policies and health economics in SA (four women and six men). It consisted of academic/researchers (n=6), government officials (n=3) and the alcohol industry (n=1).

Primary and secondary outcome measures The desktop review examined datasets for the level of the data, geographical coverage, the population surveyed, year of data collection, available covariables, analyses possible and limitations of the data. The 10 in-depth interviews with key researchers explored informant's perspective on the usability of existing alcohol datasets in SA.

Results In SA, alcohol data constraints are mainly attributed to accessibility restrictions on survey data, limited geographical coverage, lack of systematic and standardised measurement of alcohol, infrequency of surveys and the lack of transparency and public availability of industry data on production, distribution and consumption.

Conclusion The International Alcohol Control survey or a similar framework survey focusing on substance abuse should be considered for implementation at the national level. Also, alcohol research data funded by the taxpayers' money and alcohol industry data should be made publicly available.

INTRODUCTION

Alcohol abuse is a significant contributor to the global burden of disease and a cause of adverse economic impacts.^{1 2} Alcohol abuse is associated with more than 60 long-term health conditions, including cancers, cardiovascular diseases, infectious diseases and development and cognitive delays in children.^{1 3} It also affects social relationships by causing pain and suffering of family and friends of risky drinkers.⁴ Alcohol abuse is reported to reduce job productivity, increase unemployment and a drop in income levels.⁴

South Africa (SA) has particularly alarming statistics; about 7.1% of all deaths and 7.0% of total disability-adjusted life years are

Strengths and limitations of this study

- Identifies publicly available alcohol datasets and its characteristics.
- This study provides recommendations for better alcohol data collection in South Africa using key informants' experiences of dealing with alcohol datasets.
- The desktop review looked into four data warehouses.
- Informants' solutions and recommendations are based on their own experiences.

associated with alcohol consumption in the country.³ Previous studies using nationally representative survey data found that approximately half of men and one-fifth of women consume alcohol in SA and a high proportion of those who consume alcohol are likely to be involved in risky drinking.^{3 5 6} In 2015, the total per capita alcohol consumption (APC) in SA was 11.5 L of pure alcohol; while, alcohol consumption per drinker was 27 L of pure alcohol—one of the highest levels of alcohol consumption in the world.⁷

Alcohol consumption in SA is, therefore, a major problem, but many of the pathways resulting in adverse impacts are unknown. One of the challenges is that most research on alcohol consumption and its impacts in SA have had difficulty in characterising the extent and distribution at the societal level of alcohol-related harm due to data constraints. For instance, data limitations may arise from under-reporting, lack of accurate prevalence data and a failure to standardise alcohol consumption measures. A modelling study using data from informant assessments and survey data estimated that the proportion contributed by unrecorded alcohol for high-income countries was between 2.4% and 16.4% of all alcohol consumed in 2015, while it was between 9.0% and 27.6% for upper-middle-income countries. The equivalent ranges were between 38.1% and 70.8% for lower-middle-income countries and between 26.5% and 59.7% in low-income countries.⁸ Probst *et al*,⁹ using data from five nationally representative

SA surveys, estimated that the surveys only captured between 11.8% and 19.4% of total alcohol consumed per capita. That is, more than 80% of APC was unrecorded. Also, although most survey data in SA record the amount of consumption using standard drinks, the frequency of drinking recorded across different studies used different time frames.

According to Chan *et al.*,¹⁰ household surveys are the main sources of data used for monitoring and evaluating health issues, especially in low-income and middle-income countries (LMICs). Therefore, reliable and accurate household data are crucial for tracking health progress and performance and monitoring the impact of health programmes and policies.¹⁰ From 2002 to 2011, high-income countries had on average 16.8 household surveys while LMICs had, on average, 18.3 household surveys completed.¹¹ Although LMICs have more household surveys collected, most of the surveys are infrequently collected, are of poor quality or/and are not accurate.^{11 12} Glassman and Ezech¹² state that countries in Africa are in urgent need of better data. Data on poverty, births and deaths, taxes and trades, schooling and other health, economic or social welfare indicators are either missing or weak. This inadequacy impairs countries' abilities to implement efficient and effective policies.¹² For instance, Africa has the lowest coverage of birth (25%) and death (18%) registration compared with Europe, which has the highest data source on births (98%) and deaths (100%).¹³ A lack of births and deaths data make it difficult to hold governments accountable for improvements in the countries' economic and social welfare.

For alcohol, survey data can provide valuable information on patterns or problems within a community of alcohol-related harms and is commonly used by stakeholders and researchers to support policy development for alcohol control. However, the lack of systematic and standardised alcohol data may result in policies failing to address alcohol-related harms adequately. For instance, in SA, alcohol outlets have restricted opening hours laws but a high number of unregulated outlets (known as shebeens), located especially in the poorest socio-economic areas, continue to sell alcohol according to demand.⁴

To our knowledge, this paper represents the first critical assessment of the usability of data from different surveys in SA containing alcohol variables data for the assessment of alcohol consumption and related issues. This research aimed to answer the following important questions: what alcohol-related datasets are available in SA? Moreover, what are their geographical coverage, the population surveyed, year of data collection and available covariables? The research also asked what possible analysis of alcohol-related issues could be conducted using the existing datasets in SA and what are the limitations with these datasets to make recommendations for how routine datasets could be better assembled and used for informing policy? Critically, analysing the usability of alcohol data is important for, among other things,

Box 1 Eligibility criteria for inclusion used to identify South Africa alcohol datasets

1. A local, provincial or national representative survey.
2. Contains alcohol data (either consumption or expenditure or both).
3. The database is publicly available.
4. Surveys conducted after 1994*.

* postapartheid period. Most surveys prior to 1994 (during apartheid) focused more on a particular racial group (whites) and urban locations; therefore, they were not included in the analysis.

informing the effectiveness and efficacy of regulatory (eg, alcohol tax) and programmatic (eg, public awareness) interventions.

METHODS

A theory-generating, two-step qualitative methodology was used for this study. First, a desktop review of existing datasets was undertaken by the first author (MFM) to identify whether and how alcohol data were recorded. These datasets were then further examined to gain an understanding of the gaps in the literature in terms of documenting alcohol survey data in SA. Then, key researchers in alcohol policies and economics in SA were interviewed by the first author (MFM) using a semistructured, in-depth interview guide to gain the participants' perspective on the usability of existing alcohol data in SA (online supplementary appendix 1).

Desk-top review

The desktop review consisted of identifying and reviewing SA datasets containing alcohol variables to assess the usability of each of the datasets in SA. The databases were identified by accessing datasets housed in data warehouses using the eligibility criteria for inclusion contained in **box 1**. Four data warehouses (DataFirst, the International Household Survey Network (IHSN), WHO Central Data Catalog and World Bank Central Microdata Catalog), which provide a comprehensive listing of available data, were used to search for eligible datasets.

Eligibility criteria for inclusion of a dataset are listed in **box 1**. The search filters are as follows; first 'alcohol' was used as a keyword in 'variable description'; 1994 was used in 'show studies conducted' to display only studies conducted after 1994. Filter by 'data access' options 'public use data files' and 'Data available from external repository' was used to display only possible publicly available datasets and 'SA' was selected on the 'country' filter. The initial search identified 38 potentially databases using DataFirst; 52 using the IHSN; 5 using WHO Central Data Catalog and 35 using the World Bank Central Microdata. A cross-examination was performed to eliminate duplicates and datasets that were available from other sources but was not publicly available. Also, during the search, surveys with more than one round (eg, Income and Expenditure Survey 2005 and 2011) were counted as

Box 2 Eligible datasets review

- ▶ Source—provides information where the dataset is housed.
- ▶ Level of data (household or individual)—provides information on the household and/or individual level.
- ▶ Geographical coverage—describes the locations covered by the survey.
- ▶ Universe—the population that is being surveyed.
- ▶ Survey year—the year that survey was conducted.
- ▶ Measures of alcohol consumption and/or expenditure.
- ▶ Scope—covariables available (eg, demographics, socioeconomic status, health outcomes and other).
- ▶ Type of analysis that can be done using alcohol information:
 - Pricing and expenditure (topics related to pricing, eg, determine alcohol pricing, price elasticity, alcohol tax or alcohol expenditure analysis).
 - Marketing of alcoholic beverages (topics related to marketing, eg, advertising, increasing in marketing share).
 - Availability of alcohol (to track and/or reduce/increase alcohol availability, eg, restriction on alcohol sale, places where alcohol is sold).
 - The burden of alcohol (topics related to harmful use of alcohol, harm reduction, alcohol-related diseases).
 - Tracking informal alcohol consumption or sale.
 - Other (specify).
- ▶ Limitations— limitation of the survey.

one main survey (or dataset). A total of 23 datasets were identified.

The eligible datasets identified were reviewed using the categories in [box 2](#). The formulation of these categories in [box 2](#) was based on surveys description documentation and categories most likely to be used in describing a dataset in an epidemiology alcohol research.

Key informant interviews

Recruitment processes and target population

The study population was key researchers in alcohol policies and related issues in SA. We identified all researchers in SA who have done work on alcohol-related research and policy from published papers, policies and legislature documents and their work with alcohol organisations. All suitable/eligible individuals were invited to partake in the study. We specifically focused on those who conduct alcohol-related research for government, academic/research institutions, non-government and community-based organisations (NGOs/CBOs) and the alcohol industry. Inclusion criterion to identify informants required that participants have either published research on alcohol or are associated with an organisation engaged in addressing the burden of alcohol, for example, NGOs/CBOs as well as the alcohol industry. Informants were approached via email to participate in the study and were enrolled after providing informed consent.

Patient and public involvement

Study participants had no involvement in the study design or conduct of the study. The findings from this study will be disseminated to the participants via email.

Data collection process

The semistructured interviews were done face to face and in one case via Skype. Data were collected using a structured interview guide (see online supplementary appendix 1 and lasted between 30 and 60 min. It explored the informants' professional background and their experiences of dealing with alcohol datasets in SA. The interview guide was supplied to participants before the interview due to concerns about the level of information sharing. Interviews were audiorecorded and transcribed verbatim to facilitate qualitative analysis. The main questions contained in the interview guide sought to explore (1) which alcohol datasets the respondent usually used in their research, (2) which datasets they know of but have not used and (3) the reasons for not using these other datasets. Questions were also included relating to the challenges in using the datasets for research and exploring any recommendations for how routine datasets could be better used for informing policy.

The informant participants were asked to name all the datasets that they know in SA that contain alcohol-related data. The datasets provided by the informants were displayed by the themes: (1) most cited, (2) most cited but have not been used and (3) accessibility score. Parts of questions 4- 'Do you know any dataset/s* that contains alcohol related data' (where to find it? Any restrictions? Have you used it (Y/N)? Why Not?) and 5- 'Have you ever used a national or provincial dataset/s* that contains alcohol related data?' (Where to find it? Any restrictions?) from the questionnaire were used to compute the data accessibility scores. The scores were assessed on a scale of 1–5 where 1 signified most inaccessible (includes data no longer available or owners, funders or depositors of the data do not share it even if you apply for it or you only have access to the reports); 2- less accessible—(includes data that are available through reports); 3—somewhat accessible (includes data for which you must complete a form for authorisation or to request the data from the owner); 4—accessible (includes data for which you complete a form but do not need to wait for approval); 5—most accessible (data that can be accessed online without authorisation being requested). In addition, datasets cited by key informants were overlapped with the desktop review and assessed using 'Yes' if the dataset was displayed in the desktop results and 'no' if the dataset was not displayed in the desktop results.

Data analysis

All interviews were first transcribed and moved onto an Excel spreadsheet. Closed-ended questions were quantified (eg, SA alcohol datasets identified by experts) while open-ended questions (eg, What would be the possible solutions and recommendations for better alcohol data collection in SA?) were analysed manually using thematic analysis. Once the data were coded, the spreadsheet was sent to each coauthor to validate the results. Differences were discussed in the team and adjusted after reaching consensus.

RESULTS

Desktop review

A survey dataset reference list was created by identifying existing alcohol survey datasets in SA (online supplementary appendix 2). A total of 23 survey datasets were identified using the eligibility criteria (box 1). Thirteen out of 23 datasets are more than 10 years old. All the datasets addressed the burden of alcohol in some way. Eleven surveys addressed the burden of alcohol at the national level, while 12 had either municipal or provincial level. For the measure of alcohol, 20/23 surveys have individual level data on alcohol; for instance, alcohol consumption volume and frequency, safety and crime and health alcohol data. Seven out of the 23 surveys have household level alcohol data (eg, alcohol expenditure, alcohol abuse in the household and neighbourhood) and 3/23 surveys have data on alcohol at the community level (eg, crime related to alcohol and number of establishments that sells alcohol). Most commonly, survey limitations include limited national coverage, infrequent data collection intervals and surveys not collecting data needed for epidemiology research. Specifically, alcohol volume and frequency data were missing in 12 surveys or when the surveys do provide the information, the time frame of alcohol consumption or the frequency was not provided. Also, alcohol expenditure data were almost nonexistent in many surveys in SA (online supplementary appendix 2).

Themes emerging from key informant interviews

Participants

The profile of key informants is summarised in table 1. In total, 16 key informants were invited to participate in the study but only 10 (4 women and 6 men) agreed to participate (63% participation rate). It consisted of academic/researchers (n=6), government officials (n=3) and the alcohol industry (n=1). The diversity of the informants

enabled an in-depth exploration of possible solutions and recommendations for better alcohol data in SA.

Datasets cited by key informants

Table 2 shows that key informants were able to identify 24 datasets that contain alcohol data. All key informants reported use and/or had knowledge of at least one dataset. South African Demographic and Health Survey (SADHS) and National Income Dynamics Study (NIDS) were the most commonly cited datasets (n=7); however, four in seven informants have not used SADHS for analysis citing reasons such as (1) accessibility restrictions (eg, the Youth Risk Behaviour Survey has not been used because the owners, funders or depositors of the data does not share the data) and (2) the dataset contains variables that are extraneous to the informant's research interests. Only 5/24 datasets were considered most accessible and accessible (n=5 and 4); while 5/24 were somewhat accessible (n=3); 7/24 were less accessible (n=2) and 7/24 were considered inaccessible (n=1). For the top five most cited datasets, informants were more likely to use alcohol-related data for research on the burden of alcohol (topics related to harmful use of alcohol, harm reduction and alcohol-related diseases), followed by alcohol price and expenditure research (topics related to pricing, eg, to determine alcohol prices, price elasticity, alcohol tax or alcohol expenditure analysis).

In total, 6/24 datasets cited by key informants overlapped with the desktop review. The low dataset overlap was related to dataset accessibility (table 2—accessibility scores below 4). Eighteen datasets (n<4) cited by the key informants that were not identified by the desktop review are not publicly available data and can only be accessed through reports and/or are licensed data files which need authorisation from the owners, funders or depositors. Global Information System on Alcohol and Health (GISAH) was the only dataset that was identified by the

Table 1 Key informants' characteristics

Informant	Gender	Industry/sector	Primary role
Informant 01	Male	Academic/research	Researcher
Informant 02	Male	Academic/research	Researcher and manager
Informant 03	Female	Academic/research	Researcher and student
Informant 04	Male	Government	Policy-maker
Informant 05	Male	Academic/research	Researcher
Informant 06	Female	Academic/research	Researcher
Informant 07	Female	Academic/research	Researcher
Informant 08	Male	Industry	Manager
Informant 09	Male	Government	Policy-maker
Informant 10	Female	Government	Policy-maker

Four NGOs/CBOs were also invited to participate in the study but they either declined or had not replied by the time the study closed. Although no NGOs/CBOs directly participated in the analysis, some key informants work closely with NGOs/CBOs. CBO, community-based organisation; NGO, non-government organisation.

Table 2 Alcohol datasets in SA identified by key informants (n=10)

	Datasets	Dataset most cited (total)	Dataset cited but not used	Accessibility score	Overlap with the desktop review
1	South African Demographic and Health Survey	7	4	2	Yes
2	National Income Dynamics Study	7		5	Yes
3	SAARF's All Media and Products Survey	5	1	3	No
4	South African National Youth Risk Behaviour Survey	4	2	1	No
5	South African Community Epidemiology Network on Drug Use	3		2	No
6	The International Alcohol Control Study (the IAC Study)—For SA—Pretoria	3	1	3	No
7	South African National HIV Prevalence, HIV Incidence, Behaviour and Communication Survey	3		4	Yes
8	Income and Expenditure Survey	3	2	5	Yes
9	South African National Health and Nutrition Examination Survey	2	1	5	Yes
10	Global Information System on Alcohol and Health	2	1	5	No
11	Khayelitsha Household Survey	2		3	Yes
12	South Africa Stress and Health	2		3	No
13	Department of Social Development, Western Cape Resources and Services Directory for the Reduction of Harmful Alcohol and Drug Use	2		1	No
14	National Injury Mortality Surveillance System	2		2	No
15	Fetal Alcohol Syndrome dataset	1		3	No
16	South African Wine information and Systems Data	1		2	No
17	High School Survey	1		2	No
18	IRI—Sales and Marketing, Pricing Information Data	1		1	No
19	DUNNHUMBY Shopper Data	1		1	No
20	NIELSEN Survey and Electronic Data (townships and Sheebens)	1		1	No
21	Consumer Research (Industry pays for data collection)	1		1	No
22	NIELSEN Home Panel	1		1	No
23	Western Cape Emergency Health Survey	1		2	No
24	Department of Transport and Public Works (Traffic Data)	1		2	No

The scores were assessed on a scale of 1–5 where 1 signified most inaccessible; 2-less accessible; 3-somewhat accessible; 4-accessible; 5-most accessible.
SA, South Africa.

key informants and had a ‘most accessible’ score (n=4) but does not overlap with the desktop review. The reason for that is that the GISAH is not housed in data warehouses but rather on WHO webpage¹⁴ for ‘easy and rapid access’ to alcohol indicators.

Based on the desktop review and the key informants’ interviews, the frequency and volume of alcohol consumption was the variable most commonly found across available datasets; while blood alcohol concentration, alcohol price, alcohol production and purchases for firm-level variables were the least commonly variables available in datasets. For non-alcohol variables, lower level geographical coverage (eg, suburbs and townships) was generally not available, limiting the potential usability of the datasets.

Key informants’ feedback on alcohol data collection in SA

Key informants’ views on alcohol data constraints in SA showed a high degree of consensus. The major constraints are presented in four categories : (1) alcohol consumption, (2) representative alcohol data (eg, substance abuse), (3) time period/ periodicity/ frequency and (4) public availability of data on production, distribution and consumption of alcohol.

Alcohol consumption

One of the major problems in collecting alcohol data is that questions on current alcohol consumption included in many surveys generally fail to capture the true extent of alcohol consumption. According to the informants, in most surveys, there is under-reporting of alcohol consumption.

People very significantly underreport alcohol consumption and prevalence and that is a problem (Informant 02- Researcher and Manager).

One informant suggested that the survey ‘under-reporting was massive. (That is) for every four drinks a person would have they would report about 1’ (Informant 03- Researcher and Student).

Informants suggested that the reasons for alcohol being under-reported in surveys could be due to stigma, how the alcohol questions are framed, or simply because people do not know their alcohol intake levels.

One informant suggested that

... stigma—possibly have to deal with the population group of the interviewer and the gender of the interviewer. So, you can have power imbalances in the collection of the data (Informant 02 - Researcher and Manager).

While another informant suggested that

... for some reasons, people who drink alcohol seem to not quite face up to what they are drinking and also they might not be realizing how much they had. Like, if I say I had a glass and a half; I would probably say I had one glass, or I might be sitting on the interview saying oh I don’t drink at all, never drink. (Informant 03- Researcher and Student).

In addition to the alcohol questions in SA, surveys struggled to report the correct frequency and volume of alcohol consumption. Another informant pointed out the need to collect data on specific liquor types such as sugar fermented beverages, beer, wine and spirits and not by categories of harms,

At the moment, they collect separate excise taxes on beer on wine and spirits, but the definition of some of these drinks has to be improved. So, for example, the big problem is in the sugar fermented beverages (SFB). And those are really the cheapest alcohol made in this country (Informant 09- Policymaker).

No matter what the reason for under-reporting, one informant noted that

we need to get a real understanding on how much alcohol, how much standard drinks are in those things. So, is it 3 drinks or it is 3 vintage [indistinct] which are 340 mL bottles that are 2% each or is it 3 black label quarts which are 750 mL at 5.5%. So, understanding pure alcohol content across all these instruments is really important (Informant 05- Researcher).

A possible solution to overcome under-reporting suggested was the use of the

graduate frequency where you ask how often do you drink at this quantity? And also have relatively small time frames and explicit time frames to refer to. Definitely have at least quantity, frequency plus have heavy episode drinking and also possibly not just 5+drinks but 5+, 7+or like multiple categories or ask how many drinks do you drink on average on a heavy drinking occasion or like have a better assessment around heavy episodic drinking [...] (Informant 06- Researcher).

In addition, to avoid stigma an informant suggested that matching interviewers to local demographics might improve the quality of data. Another informant commented on privacy as an issue for disclosing sensitive information such as alcohol consumption. ‘I think [in] the informal housing area sometimes they don’t have a private place to have a conversation’ (Informant 06- Researcher).

Absence of a dedicated national survey for substance-related disorders including alcohol

Another alcohol data constraint cited by the informants is that there is no national dataset focusing specifically on substance abuse, especially on alcohol. According to one informant,

we need studies that look just at substance use, not as part of a survey looking on everything because you get terrible data. You need more dedicated survey looking at alcohol and other drugs use (Informant 01- Researcher).

The lack of detailed and good quality data on substance abuse may negatively influence alcohol policy interventions as, without having a proper understanding of people's alcohol consumption risky behaviours, policy implementation is likely to be ineffective. One of the informants stated that

I don't have a strong sense how to proceed on these program as a policymaker and funder. I got a perception that there is a massive substance abuse problem. I mean in the Western Cape there is a massive substance abuse problem both alcohol and drugs. I had very few requests from other departments, from the department of health or social services to say look here is a big problem we need to address in the following way etc. etc. You know, I am a bit puzzled, but they are not giving more attention to it because it is such a massive problem. There [is] a lot [of] missing information missing for supply and demand of treatment and gap. Very little is done on the control of alcohol (Informant 04- Policymaker).

Time period/periodicity/frequency

Informants mentioned that there is a need for more frequent alcohol data availability.

I am not seeing regular data coming out [...] to say that you know the number of cases of alcohol-related problems is on the regular bases. What happens with those trends and so on? So, that for me is missing. Not missing but weak. The biggest problem I am finding as we move towards to National Health Insurance [NHI] is that we have a lot of difficulties getting the Department of Health to work with [us] on the information system that is necessary (Informant 04- Policymaker).

Not having recent data available, policy-makers and researchers would not be able to provide support to advocate for policy interventions. For instance, an informant suggested that although there is data on alcohol burden, the data are not timeous.

the problem with them being these big gaps [...] it's that by the time you get numbers reported it's 2 years later and the situation could have changed. And it also does not help for planning. You know, you cannot plan. You are not working with real-time information. What ideally, we would like. I know at the moment we are trying to do this, but we need data that are more real-time from the industry and from departments that deal with social repercussion of health issues (Informant 10- Policymaker).

Public availability of data on the production, distribution and consumption of alcohol

According to the informants, there is a massive need for publicly available data on production, distribution and consumption of alcohol. Given there is a correlation

between alcohol price and consumption,¹⁵⁻¹⁷ without pricing data one cannot analyse the impact of regulatory alcohol policies that aim to reduce the affordability of alcohol beverages to decrease alcohol consumption and alcohol-related harms. One informant argued that

we need information on distribution numbers and manufacturer numbers. The distribution which you know all [the] way down to where it is delivered to local pieces that you can almost track and trace. You [are] not going to be able to track and trace but you know where the final point of the arrival is. You need figures on sales data. You need figures on pricing de-segregated by area. So, for example, at the moment it seems that the industry is adjusting prices based on the community. So, effectively what it seems to be happening is that in more dense population, in more poor areas they [industry] are selling at the lower price because the gain is that they will be pushing volume and that is how they're going to make their money as supposed to selling at a slightly higher places in more areas where there is less population but you are still going to get your targets because they will be able to afford it (Informant 10- Policymaker).

A major obstacle for alcohol-related harms research mentioned by the informants is the lack of data on the price of alcohol. According to an informant,

quantity and prices are another key thing as well. The big barrier to prices are a massive drive of consumption. So, it's pointless knowing what volume of alcohol contain beverages are sold. We need to know how much alcohol is being sold and what prices are being sold at. Because you can put so many sorts of things in place but if the price per unit of pure alcohol is decreasing, your alcohol consumption and alcohol problem will increase. Accurate price per standard drink is crucial [...] If the price of alcohol is decreasing, I guarantee the problems will increase (Informant 05- Researcher).

Possible solutions and recommendations for better alcohol data in SA good data

When asked for their opinion on what would be a perfect dataset and examples of good alcohol datasets, informants provided the following responses. One informant suggested that, overall, a good alcohol dataset should be

representative of the population sampled; clean; regularly updated; reliable and relevant to the study of interest (Informant 07- Researcher).

Another key informant agreed that a good dataset needed 'to be representative, especially Township representative' (Informant 08- Industry). As examples of good existing alcohol datasets, informants suggested WHO STEPwise approach to Surveillance (STEPS) and the International Alcohol Control Study (IAC). The STEPS

survey was described by one informant as ‘a non-communicable disease risk factor survey but includes questions on alcohol which is quite good’ (Informant 01- Researcher).

Another informant suggested that it

would be nice to have something like [the IAC survey] that is very alcohol-focused and not just [a] sideshow within the bigger survey (Informant 02- Researcher and Manager).

What should the government do to collect better data

There was not a clear consensus among the informants on what the SA government should do to collect better data. The overall comments were that the government should have a clear understanding of the data needs and find proper funding to undertake data collection. It was mentioned that even when the collection of data is funded by the government, which would normally imply that these data should be publicly available, the investigators only release it after a long delay and, as a result, the data might not be as useful for research analysis. For example, one informant referred to the South African National Health and Nutrition Examination Survey as funded by the government for which data were collected in 2011 but only released publicly in 2018. The informant noted

that is my tax money that has been used to buy data which I cannot use as a researcher. So, I think whatever the government funds, must be made public straight away. So, I think [there] needs [to be] a very open policy... (Informant 03- Researcher and Student).

In terms of alcohol industry data, it was argued that the government should enforce the public and transparent release of data on the distribution, manufacture and consumption held by the industry. One informant suggested that relying on legislation for Promoting Access in Information would not work well; rather there should be

a legislative requirement on them [industry] to provide data. Because I don't think you're going to get through. I mean you could get [data] through applying for an application [PAIA]. But that means every time you have to get information, you have to go through a court channel as if there is a legislative requirement could be a little more ongoing and transparent. And in terms of that, we need information on Distribution numbers and Manufacturers numbers (Informant 10- Policymaker).

DISCUSSION

This study examined the usability of South African alcohol data sources by documenting the type of alcohol data available in different sources and what possible alcohol analysis could be done using these datasets. It also provides

some recommendations for how routine datasets could be better used for informing policy. The results show that there are data constraints in alcohol data in SA. Through the desktop analysis, only 23 datasets met the eligibility criteria and most of these datasets are more than 10 years old and the principal agents for these surveys have now stopped collecting new data. Key informants identified 24 datasets that contain alcohol data, and 6 of them overlapped with the desktop review.

The minimal overlap between the data from the key informants and the desktop review has to do with accessibility. In the results, only 5 of the 24 datasets identified by the informants were considered ‘most accessible’ or ‘accessible’. Accessibility restrictions to alcohol datasets pose a threat to new research and the replicability of findings.¹⁸ For alcohol intervention programme and policy to be effective, they should be based on evidence-based components.^{19 20} As governments are accountable for implementing evidence-based alcohol policy, a lack of data accessibility could potentially impact the implementation of relevant policy and programme aiming to address alcohol-related harms.²⁰

A systematic review looking at the association between socioeconomic status and alcohol consumption within LMICs suggested that African surveys that collect alcohol data are ‘complicated by small non-representative samples, weak methodologies and non-significant findings’.²¹ However, none of the datasets included in the systematic review²¹ was from SA. Different from the findings by Allen *et al*,²¹ this study suggests that the constraints affecting alcohol datasets in SA are related to access restrictions to survey data, lack of systematic and standardised measurement of alcohol, limited geographical coverage, infrequent survey timing and lack of public availability of industry data on price, production, distribution and consumption of alcohol. This difference in findings may be related to political economy challenges faced by each African country. Glassman and Ezeh¹² suggested that the main challenges of data collection and use in Africa are related to offices responsible for statistics not having autonomy and stable budgets to collect data; thus, they are likely to produce unreliable and bias data. Also, donors funding projects tend to dictate how the data are collected and usually are interested in collecting micro-oriented survey and once off impact evaluations. Lastly, even when accurate data are collected, access and usability are restricted or limited.¹²

Probst *et al*⁹ confirm that alcohol consumption in SA using different nationally representative surveys is under-reported. Similar to Vellios and Van Walbeek,⁶ our results suggested that alcohol under-reporting in SA surveys might be related to the lack of systematic and standardised measurement of alcohol consumption. For instance, the NIDS Adults Survey question ‘how often respondent consumes alcohol’ might not record the actual consumption due to the absence of any time frame or recall period. In addition, an interviewee might not feel comfortable in disclosing their consumption due to

stigma. Interviewees might also have to face the challenge of not understanding the definition of standard drinks, especially low-income individuals who are more likely to consume traditional drinks such as homebrews.⁶

As SA moves towards implementing additional alcohol policies,²² it is imperative that good representative alcohol datasets are available to evaluate the effectiveness of alcohol policy interventions, among others. Therefore, this study suggests that alcohol data research in SA can be improved by making all datasets funded by the government and industry data (production, distribution and consumption data) including price data, publicly available. In addition to accessibility, substance abuse data should be collected more frequently so that policy-makers have access to 'real-time' information to evaluate and implement community evidence-based programme and policy. Lastly, it is vital to develop and test a standard alcohol questionnaire guideline for SA to be used in a national survey similar to that reported by Roche *et al*²³ which includes WHO graduated quantity-frequency measures.²⁴ The ICA dataset was cited in our results as an example of a good dataset that could potentially be a framework to provide more accurate, unbiased and consistent alcohol data in SA. Its approach to measure consumption, which accommodates country-specific beverages, was able to collect 90% of APC. Also, the IAC study provides a wide variety of relevant alcohol variables such as the frequency of drinking, typical occasional volume, quantity and alcohol purchase behaviour.²⁵ Implementation of the ICA survey at the national level may be a way to support better evidence-based alcohol programme and policy.

We believe that SA's experience may be quite different from other LMICs with different research and surveillance environments. However, for countries wishing to revamp or improve its collection of national alcohol data sources, we suggest the following steps in assessing the usability of alcohol datasets: (1) document all the datasets that exist, (2) consider measures to ensure public availability of data, (3) try to harmonise key measures (eg, how to measure alcohol consumption, how to measure alcohol spending, time periods linked to both) while allowing diversity in other variables collected and (4) include a measure of the quality of the data.

LIMITATION AND STRENGTH

One of the study's limitations is that the desktop review only looked into four data warehouses; however, these warehouses provide a comprehensive listing of many datasets conducted in SA. Key informants did not include medical professionals who might have a good insight on alcohol datasets. Another limitation is that informants' solutions and recommendations are based on their own experiences, making them vulnerable to bias. Nevertheless, they are stakeholders and have a good understanding of the data constraints. The advantage of using in-depth-interviews with key informant was that it

enabled the identification of the alcohol datasets used by informants and their uses. These datasets overlapped with the publicly available data from the desktop review. This study was also able to provide recommendations for better alcohol data collection in SA using key informants' experiences of dealing with alcohol datasets. It also shows publicly available data and their characteristics.

CONCLUSION

Alcohol policy and programme interventions are more likely to have a more significant impact on decreasing harms when they are based on evidence. Based on the findings of this study, it is suggested that the ICA survey or a similar framework survey focusing on substance abuse may be considered for implementation on the national level. Also, alcohol data funded by the government and industry data should be made available to the public. It is by having accessible, reliable and meaningful data that stakeholders and researchers can evaluate interventions.

Acknowledgements JA is supported by the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation.

Contributors MFM carried out the desktop review, interviews with key informants and led the writing of the manuscript. MFM, LL, NHB and JA helped to conceptualise the research, reviewed the results, helped to write and revise the manuscript, and approved the manuscript submitted for publication.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The study was approved by the Human Research Ethics Committee of the Faculty of Health Sciences at the University of Cape Town (HREC reference number: 798/2017).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement The data that support the findings of this study are available on request from the corresponding author, MFM and will be required to agree to the Terms and Conditions of a Data Access Agreement (DAA), which aims to protect the privacy and interests of the research participants.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

1. Rehm J, Room R, Monteiro M, *et al*. Alcohol as a risk factor for global burden of disease. *Eur Addict Res* 2003;9:157–64.
2. Rehm J, Mathers C, Popova S, *et al*. Global burden of disease and injury and economic cost attributable to alcohol use and alcohol-use disorders. *Lancet* 2009;373:2223–33.
3. Schneider M, Norman R, Parry C, *et al*. Estimating the burden of disease attributable to alcohol use in South Africa in 2000. *S Afr Med J* 2007;97:664–72.
4. Setlalentoa B, Pisa P, Thekisho G, *et al*. The social aspects of alcohol misuse/abuse in South Africa. *South African Journal of Clinical Nutrition* 2010;23:11–15.
5. Parry CDH, Plüddemann A, Steyn K, *et al*. Alcohol use in South Africa: findings from the first demographic and health survey (1998). *J Stud Alcohol* 2005;66:91–7.

6. Vellios NG, Van Walbeek CP. Self-reported alcohol use and binge drinking in South Africa: evidence from the national income dynamics study, 2014 - 2015. *S Afr Med J* 2018;108:33–9.
7. World Health Organization. Global health Observatory data Repository, 2016. Available: <http://apps.who.int/gho/data/view.main.GDO2103v?lang=en> [Accessed Nov 2018].
8. Probst C, Manthey J, Merey A, *et al*. Unrecorded alcohol use: a global modelling study based on nominal group assessments and survey data. *Addiction* 2018;113:1231–41.
9. Probst C, Shuper PA, Rehm J. Coverage of alcohol consumption by national surveys in South Africa. *Addiction* 2017;112:705–10.
10. Chan M, Kazatchkine M, Lob-Levyt J, *et al*. Meeting the demand for results and accountability: a call for action on health data from eight global health agencies. *PLoS Med* 2010;7:e1000223.
11. Serajuddin U, Uematsu H, Wieser C, *et al*. *Data deprivation: another deprivation to end*. The World Bank, 2015.
12. Glassman A, Ezeh A. *Delivering on the data revolution in Sub-Saharan Africa*. Washington, DC: Center for Global Development, 2014.
13. United Nations. Coverage of birth and death registration, 2016. Available: <https://unstats.un.org/unsd/demographic-social/crvs/index.cshhtml>
14. World Health Organization. Global information system on alcohol and health (GISAH). Available: <http://apps.who.int/gho/data/node.gisah.GISAHhome?showonly=GISAH2019> [Accessed May 2019].
15. Chaloupka FJ, Grossman M, Saffer H. The effects of price on alcohol consumption and alcohol-related problems. *Alcohol Res Health* 2002;26:22–34.
16. Xu X, Chaloupka FJ. The effects of prices on alcohol use and its consequences. *Alcohol Res Health* 2011;34.
17. Elder RW, Lawrence B, Ferguson A, *et al*. The effectiveness of Tax policy interventions for reducing excessive alcohol consumption and related harms. *Am J Prev Med* 2010;38:217–29.
18. Anderson BA. The role of data in research and policy. In: *Emerging trends in the social and behavioral sciences: an interdisciplinary, searchable, and Linkable resource*, 2015: 1–13.
19. World Health Organization. *Evidence for the effectiveness and cost-effectiveness of interventions to reduce alcohol-related harm*. World Health Organization, 2009.
20. US Department of Health Human Services. *Facing addiction in America: The Surgeon General's report on alcohol, drugs, and health*. US Department of Health Human Services, 2016.
21. Allen LN, Townsend N, Williams J, *et al*. Socioeconomic status and alcohol use in low- and lower-middle income countries: a systematic review. *Alcohol* 2018;70:23–31.
22. Parliament of the Republic of SA. *Liquor Amendment Bill, 2018 - Notice of intention to introduce a Private Members Bill into Parliament and invitation for public comment thereon National Gazette, No. 41957*, 2018.
23. Roche A, Pidd K, Taylor C. *Alcohol data reporting standards*. Adelaide: National Centre for Education and Training on Addiction (NCETA). Flinders University, 2011.
24. World Health Organization. *International guide for monitoring alcohol consumption and related harm*, 2000.
25. Huckle T, Casswell S, Mackintosh AM, *et al*. The International alcohol control study: methodology and implementation. *Drug Alcohol Rev* 2018;37(Suppl 2):S10–17.