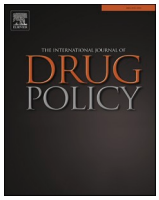
International Journal of Drug Policy 148 (2026) 105115

Contents lists available at ScienceDirect

International Journal of Drug Policy

journal homepage: www.elsevier.com/locate/drugpo



Leveraging web-scraped data to examine alcohol pricing: an Australian feasibility study with retail data

Tina Lam [a,*], Cynthia Huang [b], Alexandra Torney [c], Sarah Callinan [c], Brian Vandenberg [a,d], Ting Xia [a], Colin Angus [e], Robin Room [c], Rowan Ogeil [a,f], Aaron Cowper [g], Simone Pettigrew [h], Bosco Rowland [a,f], Danica Keric [i], Dan I Lubman [a,f], Suzanne Nielsen [a,f]

[a] Monash Addiction Research Centre, Monash University, Victoria, Australia
[b] Department of Econometrics and Business Statistics, Monash University, Victoria, Australia
[c] Centre for Alcohol Policy Research, La Trobe University, Victoria, Australia
[d] Productivity Commission, Australian Government, Australia
[e] The Sheffield Addictions Research Group, University of Sheffield, England, UK
[f] Turning Point, Eastern Health, Victoria, Australia
[g] ShopGrok, Chippendale, New South Wales, Australia
[h] The George Institute for Global Health, University of New South Wales Sydney, New South Wales, Australia
[i] Cancer Council Western Australia, Western Australia, Australia

ARTICLE INFO

Keywords:
Alcohol
Alcohol price
Alcohol policy
Pricing
Health economics
Public health policy
Online surveillance

ABSTRACT

Introduction: Although price is critical in determining alcohol purchase and subsequent harms, researchers rarely have access to comprehensive alcohol price data. Web scraping is an advanced data collection technique that uses automated computer scripts to efficiently gather extensive website data. The aims of this paper are to demonstrate web scraping's capacity to generate alcohol policy relevant data, and to assess the method's consistency by comparing datasets collected by a commercial provider with those produced by a university-developed scraper.

Methods: Price and product data from the entire online catalogues of major retailers representing the majority of the Australian market were scraped daily by the commercial provider since 2020, with data collected from all jurisdictions, and products sold by multiple retailers matched. A university-developed web scraper collected a single-day's catalogue data from the country's largest alcohol retailer to compare with the commercial dataset as a reliability cross-check.

Results: Of the 16,409 products identified in both the commercial and university databases, there was an excellent match on the product prices (intraclass correlation coefficient=0.997 [95 %CI: 0.9972–0.9973]). A visualisation from the three largest Australian retailers demonstrated how daily prices varied over a 12-month period, for example with more frequent price changes for Australia's largest retailer compared to the second and third, and across jurisdictions, such as some deeper discounting in Victoria.

Discussion: This study presented an independently cross-checked large-scale and longitudinal web scraping approach to collect alcohol price data, and demonstrated that the adapted data could aid understanding of the alcohol retail market. Web scraping is a feasible method to collect price data to support the development of evidence-based alcohol price policy.

Introduction

Globally, alcohol causes 2.6 million deaths per year (World Health Organization, 2024), and is a top five risk factor for disease burden for 15–49 year olds (Brauer et al., 2024). In Australia, alcohol is a driver of 250,000 emergency department presentations each year, family

Given their role as Associate Editor, Sarah Callinan had no involvement in the peer-review of this article and has no access to information regarding its peer-review. Full responsibility for the editorial process for this article was delegated to another journal editor.

* Corresponding author.
E-mail address: tina.lam@monash.edu (T. Lam).

https://doi.org/10.1016/j.drugpo.2025.105115

Available online 20 December 2025

violence, as well as longer-term harms, such as 1 in 20 cancers (Whetton et al., 2021). The annual cost to Australian communities of these harms to is conservatively estimated at $67 billion (Whetton et al., 2021), with alcohol tax revenue recouping only a tenth of this cost (Australian Government, 2018).

The vast majority of alcohol consumed in Australia is purchased through licensed venues for either off or on-premise consumption. It is estimated that around 80 % of alcohol is purchased 'off-premise' (i.e. takeaway) from retail bottle stores (Jiang, Callinan, Livingston & Room, 2017). Compared to other outlet types (e.g. bars and restaurants), off-premise outlets have a much greater impact on harms such as assault, traffic crashes, and family violence, explained in part as due to the much lower prices available in the packaged liquor market (Morrison et al., 2016).

Price matters in purchasing

There is strong evidence that price is a key driver of behaviour (Sharma, Sinha & Vandenberg, 2017). It has long been established that people who drink at higher levels tend to consume higher volumes of the cheapest alcohol which in turn contributes disproportionately to alcohol-related harm (Callinan, Room, Livingston & Jiang, 2015; Gill et al., 2015). To reduce alcohol-associated harms, the World Health Organization considers pricing policies, such as raising taxes, among the most effective interventions that can reduce the most harm at the lowest cost, in both general and high-risk populations, and specifically recommends the ongoing monitoring of alcohol prices (World Health Organization, 2019).

Even a small increase in price can have a substantial impact on demand. Meta-analyses and reviews find a 1 % increase in price will on average reduce demand by 0.3–0.77 % (Sharma et al., 2017), with an associated reduction in harms (Secombe, Campbell, Brown, Bailey & Pilcher, 2021). A 10 % increase in the average price of the cheapest alcohol has been associated with a 32 % decrease in alcohol attributable deaths (Zhao et al., 2013).

Better understanding pricing can be used to develop and implement more effective public health price policies and interventions. Researchers are not only interested in typical alcohol prices, but in the distribution of prices, for example, due to the greater impact of cheaper alcohol on harms. However, there is remarkably little data on the price of alcohol, with most research in the field relying on purchaser recollection of what they paid for their alcohol (Callinan et al., 2016, 2015; Rehm, Kilian, Rovira, Shield & Manthey, 2021). Further, examinations from the UK (Seaton & Waterson, 2013), New Zealand (Bentley, Chipeniuk & Krsinich, 2024), and Australia (Australian Bureau of Statistics (ABS), 2019) report that alcohol products have some of the most frequent pricing fluctuations of all food and beverage categories. Data collection on a frequent level is therefore necessary to provide a reliable and representative measure of volatile pricing.

Existing methods of collecting alcohol price data

'Mixed capitalist economies', found in countries like the USA, UK, and Australia, are the most common type of alcohol retail market. They combine private for-profit interests such as marketing by alcohol companies, with government regulations such as an alcohol retail licensing system and alcohol taxes (Economics in Context Initiative, 2021). In contrast, the most wide-ranging overviews of prices to date are from locations such as Sweden and certain Canadian provinces, where much of the alcohol retail sector is operated as a government monopoly, with strict control on promotions and without private profit incentive (Gruenewald, Ponicki, Holder & Romelsjö, 2006; Stockwell, Churchill, Sherk, Sorge & Gruenewald, 2020). Although a large volume of alcohol product-specific records are available from these government-owned liquor chains, the results of studies conducted in these jurisdictions have limited applicability to mixed economies. This lack of comprehensive alcohol pricing data stymies the development and implementation of evidence-based pricing policies in mixed economies (Fogarty & Chapman, 2013).

With the alcohol industry not required to publicly disclose detailed information on alcohol retail prices or volumes sold, research on retail prices has typically utilised a mix of self-report and high-level sales data. These studies have largely been focused in jurisdictions where pricing policies have been or are proposed to be implemented, such as the United Kingdom (Brennan, Meng, Holmes, Hill-McManus & Meier, 2014; Holmes et al., 2014; O'Donnell et al., 2019), Canada (Hill-McManus et al., 2012), and the United States (Bertin et al., 2024). While surveys are able to link price paid to individual consumption, both expenditure and consumption are generally underestimated, response rates are at a historic low, and people who consume alcohol at risky levels are less likely to respond to surveys (Callinan et al., 2016, 2015; Rehm et al., 2021). Paid market research panels that require consumers to hand-scan the barcodes of purchased products can be very expensive, commonly report prices averaged over time (Cavallo & Rigobon, 2016), and tend to overrepresent price-sensitive participants (Lusk & Brooks, 2011). Wholesale liquor sales data are sometimes reported by suppliers for regulatory or taxation purposes, but these are only available in some jurisdictions, at an aggregate level, and do not include the product-level retail price information most relevant to individual consumers (Rankin & Livingston, 2016). While researchers' in-store (Jones, Barrie, Robinson, Allsop & Chikritzhs, 2012) or website-based surveys (Lam et al., 2023) can have a defined sampling frame, the manual recording of prices is too onerous to conduct in a comprehensive (product-specific) and frequent manner. Further information on the benefits and considerations of methods used to estimate alcohol prices are provided in Supplementary Table 1.

Using online prices and web scraping as a streamlined data-gathering technique

Web scraping is a relatively new and innovative technology used to extract data from websites via automated computer scripts (Rennie et al., 2020). Web scraped retail data can be used for a variety of purposes and has been identified as an important data collection tool by government agencies involved in the production of consumer price indexes (Australian Bureau of Statistics (ABS), 2019; Cavallo & Rigobon, 2016) and has increasingly been used to collect food and beverage information to inform public policy (Davies, Lam & Pettigrew, 2025, 2024; Hillen, 2019; Ravandi et al., 2025). Compared to the previously described price collection methods, collecting price information via web scraping is timely, systematic, less labour-intensive, far more comprehensive, and able to be deployed on a much more frequent schedule (e.g., daily). Despite these many advantages, the technical complexity and infrastructure requirements of web scraping systems are known obstacles to ongoing collection for research purposes (Albis, Romasoc, Pelayo, Gavira & Asombrado, 2023; Cavallo & Rigobon, 2016).

The aims of this paper are to assess web scraping's capacity to generate alcohol policy relevant data, and to demonstrate the method's consistency by comparing datasets collected by a commercial provider with those produced by a university-developed scraper. We investigate these aims using the following research questions:

1. What types of alcohol retail information can be collected with web-scraping methods?
2. Can commercial web-scraped information be adapted to examine the availability of different products?
3. How can we adapt commercial web-scraped information to examine price fluctuations?
4. Is commercial web-scraped alcohol price information reliable?

Methods

Specific study context and rationale

In 2021, this research team collected online prices for around 28,000 alcohol products across 16 off-premises and 11 on-premise alcohol outlets (Lam et al., 2023). To supplement this website-based but largely manually collected survey, the research team developed a custom web scraper to successfully collect information on around 19,000 products from Australia's largest alcohol retailer (Lam et al., 2021). Due to the data engineering demands, the research team's scraper was run at a relatively infrequent interval, prompting consideration of which day would best capture a representative snapshot of alcohol prices. As there was an absence of literature describing highly-specific temporal fluctuations in alcohol prices, the team identified a need for daily data to answer the fundamental research question of whether 'typical' pricing patterns existed. However, the technical demands of conducting daily scrapes were not feasible in a research setting, prompting the team to investigate the availability of commercial data products. The team learned that large retailers were faced with a similar problem of needing rapidly updated product listings, in the retailers' case, in order to price-match competitors. The retail sector's requirement to sample entire catalogues on a daily basis then prompted the research team to locate a third-party data provider who serviced this requirement with well-established commercial-level infrastructure.

ShopGrok is the largest commercial data aggregator within the Australian fast-moving consumer goods sector, serving more than a quarter of the top 20 largest retailers. They collect data from a range of sectors, including an extensive range of alcohol retailers, and importantly, collect this data on a daily basis while matching products across retailers. To establish an appropriate academic-commercial partnership, the lead investigator (TL) contacted ShopGrok (AC) to determine their interest in providing data for research purposes, and established a standard Research Collaboration Agreement For an Investigator Initiated Project (Monash Partners, 2024). The research team is solely responsible for identifying research questions of interest. This agreement allows the University to publish study results, with the proviso that 28 days prior to any publication, ShopGrok are given the opportunity to provide comments that protect its confidentiality and/or background IP. The study data reported herein is from ShopGrok unless otherwise specified.

Web scraping for data acquisition

The three main data engineering steps of web scraping are: (1) Indexing - listing all the URLs with target content, (2) Fetching content from the listed URLs, and (3) Parsing - transforming collected unstructured data into structured data (Rennie et al., 2020). A visual overview of this workflow is provided in Fig. 1. There is a spectrum of methods to achieve these steps - from the use of no-code products that can be installed as a browser widget, to options that require greater data engineering expertise, such as adapting open-source web scraping code libraries and writing custom code for the target website. The method selected is in part dictated by the complexity of the target website, with more complex target websites requiring the more resource intense custom coding options. More complex target websites include 'dynamic' content that can be customised to the visitor (such as a different price depending on the visitor's location), a high volume of content, and anti-scraping measures. One of the common technical challenges to maintaining an ongoing stream of reliable and usable data, is "scraper break" where the data collection process is disrupted due to changes on the target website structure, anti-scraping security measures, or technical errors on the computers and servers used to scrape the data (Albis et al., 2023; Khder, 2021).

Metadata overview

As alcohol prices fluctuate frequently and historical figures are not typically published by retailers, daily scraping is desirable to generate longitudinal data with products nested within retailers. The commercial partner collects prices for all products available online from major retailers across all eight Australian jurisdictions on a daily basis. The dataset included product-level information such as product and brand name, beverage category (e.g., beer, wine, spirits), price, retailer, provider-generated unique product identifiers, store location, package volume, units per pack, manufacturer, and alcohol by volume.

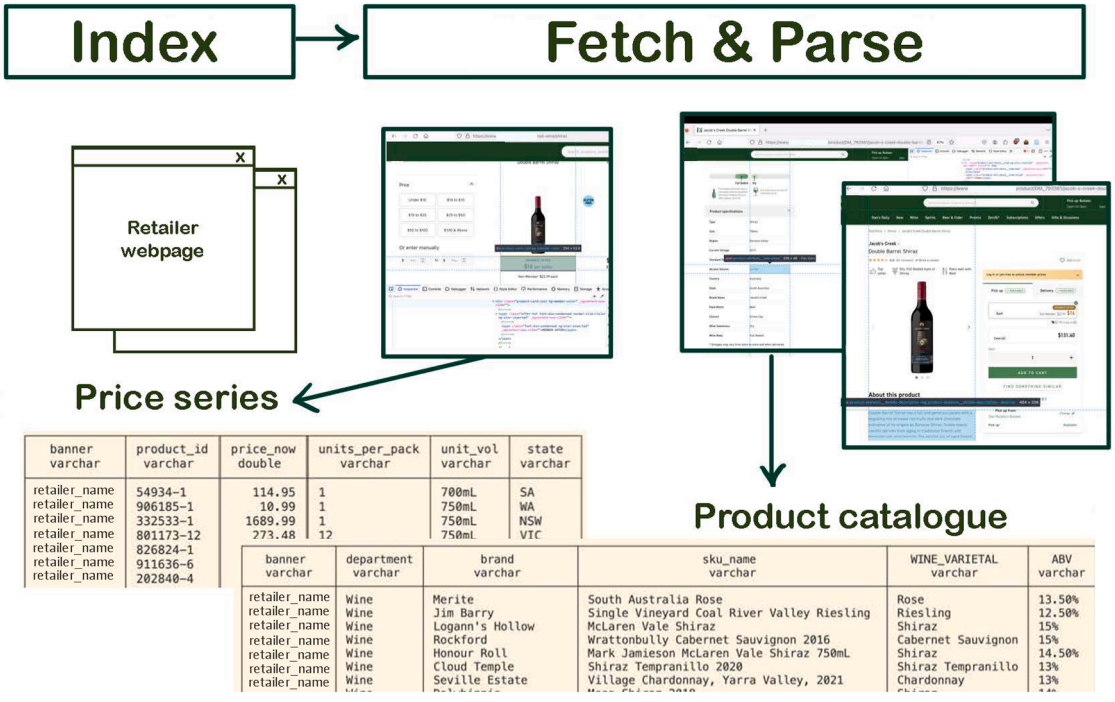


Fig. 1. Overview of the web scraping workflow used to collect alcohol price data.

Data harmonisation

Considerable variation existed in the structure and presentation of product information between retailers. Examples include inconsistent naming conventions (e.g., 'three' or '3′ in brand names) and variability in the completeness of ancillary characteristics such as region of origin. Accordingly, the commercial data aggregator's harmonisation processes were applied to standardise product attributes and align listings referring to the same product across different retailers. There are several methods that can be used for product harmonisation (with our commercial provider employing barcode matching, fuzzy product name matching, and image comparison). The majority of products are available at multiple outlets, and product harmonisation allows for the recognition of items sold by multiple outlets and retailers, as opposed to being exclusive to a single retailer.

Retailer sampling frame

The population of interest are Australian off-premise alcohol retailers, and this section outlines our sampling frame of these retailers. For the purposes of this paper, we selected a sub-sample of the data available from the commercial partner, to test the research questions concerned with product availability and pricing. While the commercial provider collects data from 39 retailer brands ($n = 35$ alcohol retailers and $n = 4$ retailers specialising in zero-alcohol beverages, since 2020), we focused our analyses for this paper on the "major" (largest) retailers which are all currently or previously supermarket affiliated and cumulatively represent the majority of the Australian retail market (see Table 1). This focus on the largest retailers was grounded in both practical and strategic considerations. From a data management perspective, incorporating additional retailers significantly increased the complexity and labour associated with data wrangling and cleaning, with diminishing returns in terms of added market coverage and policy relevance. Moreover, as public health researchers, we were particularly interested in characterising the availability of low-cost alcohol products to the largest proportion of the population - a dynamic primarily driven by the dominant supermarket-affiliated duopoly in Australia (Reeves, 2021; Roy Morgan, 2019; Roy, 2017).

The sampling unit is a retail store; that is, a bricks and mortar location of a given retailer. A "primary store" was identified for each retailer in every jurisdiction, typically from a CBD/inner city location with a wide product range. Only the primary store is sampled if the retailer has a jurisdiction-level pricing strategy (see Table 1), but multiple stores within the same jurisdiction were sampled if the retailer has greater variation in their pricing structure.

Table 1
Major Australian alcohol retailers.

Retailer	Market share[1]	Pricing structure [2]	Ownership
Retailer A	30 %	Dynamic	Australia's largest retail drinks and hospitality business. Prior to 2021, this business was a part of one of Australia's duopoly supermarket chains.
Retailer B	17 %	National/state-wide	
Retailer C	9 %	National/state-wide	Major supermarket chain two (the second of Australia's duopoly supermarket chains).
Retailer D [3]	5 %	National/state-wide	
Retailer E [3]	Not available	National/state-wide	
Retailer F	5 %	National/state-wide	European owned supermarket chain
Retailer G	5 %	Catalogue specials shared across the supermarkets	An independent supermarket brand, with individual stores owned and operated independently.

[1] Estimated Australian alcohol retail market share from Roy Morgan & IBIS World (Reeves, 2021; Roy Morgan, 2019; Roy, 2017).

[2] Hypothesised pricing structure informed by retailer policies as shown on their websites and as understood by the commercial data custodians in 2025, subject to change. Dynamic pricing is a strategy that fluctuates depending on factors such as the prices listed by geographically proximal competitors (e.g. within 10 km of their target store).

[3] In 2025, it was announced Retailer D and Retailer E stores were to be rebranded under retailer C's branding (Parkes-Hupton, 2025).

Cross-check on data quality via a university collected web scrape

The risk of discontinuity, scraper break or other issues was mitigated through daily quality assurance and monitoring by the commercial data provider, including internal checks for anomalies such as sudden or extreme price changes. There is also an external feedback channel whereby clients can flag issues such as mismatched products for investigation and correction by the commercial partner's full-time data engineering team.

We sought to assess the suitability of the commercial price database for research use, and undertook the following processes. We gained a detailed understanding of the commercial partner's data collection and quality control processes through database interrogation as well as regular meetings with the data providers. Secondly, as an original analysis for this paper, we built an 'in-house' custom web-scraper targeting the entire catalogue of Retailer A (Australia's largest retailer by market share and product range), to conduct an independent cross-check on the data points available within the commercial dataset. The university web-scrape used to cross-check the commercial data was collected on a single to minimise collection instability issues such as scraper break. The proportion of overlap between the commercial and university scrapes in product coverage and price accuracy was reported along with an intraclass correlation coefficient, a metric on the level of price-match agreement. Further information on the commercial provider's quality control procedures and the research team's independent cross-check is presented in Supplementary Box 2 and Box 3, and we describe our investigation and evaluation loops of the commercial dataset elsewhere (Huang & Lam, 2025).

Ethics

Monash University Ethics confirmed that external ethical review was not required for this study as it used existing publicly available non-human data and no human participants were involved (per NHMRC 2007 National Statement on Ethical Conduct in Human Research).

Results

Overview of information collected

The current commercial database contains daily prices for over 16,500 harmonised and retailer-exclusive products available for sale by one of the represented retailers. A summary of the study data characteristics is shown in Table 2.

Reliability cross-check

The cross-check between the university and commercial dataset for a single day of data from the largest retailer showed high correlation between the two datasets. The intraclass correlation coefficient between the two datasets (ICC [2,1]) was calculated using a two-way random effects model with absolute agreement, and was 0.997 ($n = 16,409$ products, CI: 0.9972–0.9973). ICC values above 0.9 are indicative of excellent reliability (Koo & Li, 2016). See supplementary Box 3 for

Table 2
Summary of data source market coverage and sampling schedule.

Dimension	Description
Retailers	Up to 39 relevant retailers ($n = 35$ alcohol retailers and $n = 4$ retailers specialising in zero-alcohol beverages), with the largest retailers outlined in Table 1.
Product range	Data collected on all available products (~16,000 for the largest retailer) including product name, beverage category, price, volume per unit, unit per pack (e.g. single 375 ml can, 6 pack, 24 pack, 30 pack), alcohol by volume, packaging type (e.g. can, bottle or cask), and manufacturer.
Jurisdiction	All of Australia's eight jurisdictions are sampled. The catalogues of least one store per major retailer (Retailers A, B, C, D, E and G from Table 1) per jurisdiction are sampled daily.
Sampling frequency	Major retailer stores in every jurisdiction are sampled at least once a day (i.e. some stores are sampled multiple times a day). There has been continuous daily data collection since 2020.

further information on the results of the comparison between the datasets.

Cross-sectional comparison between products

The Fig. 2 histogram offers insight into how cross-sectional data can demonstrate general pricing patterns across beverage categories. Fig. 2 shows the distribution of prices for individual packaging units (e.g. a 375 ml can of beer as opposed to a 6 pack of beer) across three beverage categories, based on data from a single day from Retailer A. Prices were averaged across all store locations for each unique product. The prices are shown in a log scale in the figure to account for the wider price range at the upper end of the distribution. Beverage categories reflect the retailer's own classification.

Temporal granularity of pricing data

The visualisation in Fig. 3 offers insight into how daily price fluctuations can be captured for the same product to compare how price can vary across jurisdictions, retailers and time. Specifically, it shows the daily fluctuations in the daily average price of a single bottle of a shiraz/syrah wine across all eight Australian jurisdictions, across the three largest retailers, May 2024 to May 2025. A single store was chosen to represent each retailer in each jurisdiction, with the store selected on the basis of having the highest data coverage for the time period. Retailer A's data is not presented for NT where there are no Retailer A stores.

Fig. 3 illustrates how prices vary over time, with more frequent price changes for the largest retailer compared to the second and third largest retailers, as well as between jurisdictions, e.g. deeper and more sustained discounting in Retailer B in Victoria than the rest of Australia.

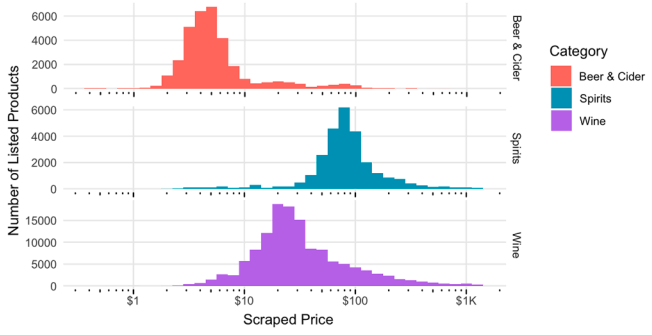


Fig. 2. Distribution of single-unit prices by beverage category on a single day from one retailer, with prices shown in a log scale.

Discussion

Alcohol pricing policies are one of the most effective strategies to reduce alcohol harm, and understanding alcohol pricing is a fundamental step in developing these policies. In recent years there has been a rapid increase in the quantity and quality of alcohol product data available online, however its collection and aggregation is not straightforward. In the absence of a government mandated retailer reporting system, web scraping is a feasible way to gain access to this price data.

In this study, we investigated the use of a commercially sourced web-scraped alcohol price dataset that provided broad retailer, product and time series coverage of the Australian alcohol retail sector. These data are routinely used for retail analytics and competitor monitoring, creating strong commercial incentives for accuracy. To further assess data quality, we compared results with our own university-run web scrape of alcohol prices. The strong consistency observed between commercial and university sources supports our proposal that both datasets are representative of a substantial portion of Australian alcohol retail market prices (Kruskal & Mosteller, 1979). We then extracted and adapted a comprehensive variety of information from the primary (commercial) dataset such as the number of products available within a beverage category, and the price of each product, to illustrate information that can be useful for policy development such as price variation across retailers, jurisdictions and time.

Translation to broader research contexts

One key area of public health research interest commonly centre in pricing changes over time, which requires more frequent data collection. However, frequent data collection from dynamic retail websites presents substantial technical challenges. Dynamic websites are prone to changes that can "break" scrapers, requiring custom code for each site as well as ongoing monitoring and recoding to maintain functionality. When sampling is required at high frequency, these demands quickly exceed the resources typically available for a time-limited research project. Similarly, there are substantial and ongoing data engineering requirements to identify and reconcile variations in retailer catalogues to reliably and accurately match products across retailers. This harmonisation of variation is useful as it allows for the observation of competitive retailer pricing dynamics - an important driver of discounting strategies that influence alcohol sales volumes and associated harms. For these reasons, the research team opted to partner with a commercial data provider, whose business model is built on sustaining the infrastructure and data engineering expertise needed to deliver reliable, large-scale, frequently updated data that can be compared across retailers.

However, it is important to note that a highly complex, daily-scraped web scraping system is not essential for addressing many policy-relevant research questions. The research question itself determines the data required, and thus whether a more complex system with substantial software developer and data engineering resources is required. For example, a one-off cross-sectional scrape of a single major retailer is within reach for many research units and can provide valuable insights into the structure of a significant portion of a region's alcohol retail market, including the range of available products, brands, products' countries of origin, and pricing variations by beverage category. Semi-regular scrapes of the same major retailer may be used to examine inflation-adjusted changes in a sample of cheaper versus more expensive products. So, while more frequent sampling enables the study of dynamic practices, such as event-driven discounting and a day-by-day capture of how retailers adjust the extent to which they pass on tax increases to customers, less frequent sampling remains adequate to answer many important public health questions.

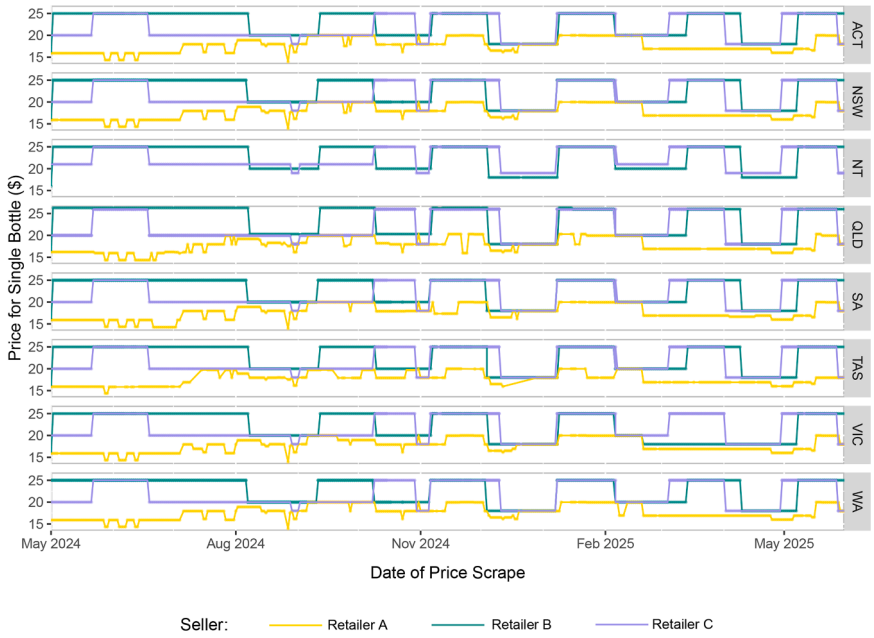


Fig. 3. Demonstration of fluctuations in the price of a bottle of wine from the three largest Australian retailers across eight jurisdictions between May 2024-May 2025. y-axis is in Australian dollars (1 AUD ≈ 0.66 USD in late 2025). ACT=Australian Capital Territory, NSW=New South Wales, NT= Northern Territory, QLD=Queensland, SA=South Australia, TAS=Tasmania, VIC=Victoria, WA=Western Australia.

Value for government and policy development

Reliable longitudinal detailed alcohol price and product data can be invaluable when used in impact analyses or policy evaluations to help policymakers consider how proposed or implemented policies affect individuals, industry, and the broader community (Office of Impact Analysis, 2023). Although all impact modelling and evaluations of price-based policies require prices as variables, Australian prices used for analyses have only previously been estimated using imputed or a small number of real prices (Srivastava, Yang & Zhao, 2022). The sheer volume and product-specific detail captured by web-scraping techniques means that modelling can use real-world, real-time, representative and reliable price data for the market. Also, the adaptation of more comprehensive price data enables more granular and accurate assessment of the connection between alcohol affordability and demand. For example, from 1980 to 2012, relative Australian wine prices fell while beer and spirit prices rose, which was associated with a change in their relative popularity over this period (Jiang and Livingston, 2015).

The inclusion of real-world detail is crucial for public communications, enabling policymakers to demonstrate that current and proposed scenarios are grounded in actual price data and accurately reflect the marketplace. This also allows them to present credible and relatable examples to enhance public understanding of policy debates. This is critical, as price impacts are often mis-estimated by alcohol retailers and the general public, undermining support for the most effective levers available to governments to reduce public health harm (Chalmers, Carragher, Davoren & O'Brien, 2013; Stockwell, Giesbrecht, Vallance & Wettlaufer, 2021).

More generally, we argue that with appropriate aggregation, online prices can be a meaningful and reliable indicator of both online and in-store prices. It has been documented there is high consistency in online and in-store prices, including within the Australian alcohol retail context (Zorbas et al., 2021), and prices have been converging over time, likely due to consumer comparison pressures and the common use of pricing guarantees policies (Cavallo & Rigobon, 2016). While most Australian alcohol sales remain in-store, with only 12 % of 2024–25 total sales being online, these online sales primarily occur through the dominant alcohol retailer duopoly (Reeves, 2025). The increasing introduction of electronic shelf labels (ESLs) that can be updated centrally rather than by individuals in-store, further decreases the likelihood of systematic discrepancies between online and offline prices (Murti, 2024). ESLs are associated with considerable cost savings for the retailer, makes it much easier for retailers to operate dynamic pricing strategies, and as of June 2024, all stores from Australia's largest alcohol retailer had ESLs installed (Profaca, 2024).

Strengths and limitations

This study used a commercially web scraped dataset which offered granular extensive temporal and jurisdictional coverage, as well as cross-retailer product harmonisation. To our knowledge, this represents the most comprehensive and organised aggregation of alcohol pricing data available in any international context outside of a government monopoly retail system. More specifically, the data illustrated in our study improves on the limited number of mixed economy studies that use web scraping to collect alcohol price data by capturing data more frequently, over a longer time period, and across multiple retailers. Previous studies typically collected data over short time frames, and with either a single retailer (e.g. Burton, Henn, Fitzgerald & Sheron, 2024), or combined all products without accounting for duplicates - resulting in double-counting when the same product appeared across multiple retailers (e.g. Davies et al., 2025, 2024).

Web scraping procedures can lead to sampling bias, in particular those emerging from the source data changing over time, being personalised and unindexed (sampled from an uncharacterised population) (Foerderer, 2023). In this study, in order to mitigate these risks to the validity of downstream analysis, we aligned the research questions and methods to correct for or minimise these biases, or to utilise these features as analysis opportunities. Firstly, the alcohol data is scraped by the commercial providers on a daily basis, so researchers are able to capture the temporal volatility and utilise it for time-series analyses. Secondly, as data is personalised to the selected store, store location becomes available for spatial analyses. Thirdly, the relatively constrained research context (the Australian alcohol retail market) meant we were able to carefully characterise the available sample via indexed units such as market share estimated for each retailer and liquor licenses owned by

each retailer to inform the retailer selection criteria.

Lastly, although the collection of public alcohol price data tells us what is available for purchase, it importantly does not indicate the volume sold or consumed. Future research could explore the impact of price fluctuations through integration with other datasets such as point of sales data or those related to alcohol related harms. Though this study only presented demonstration data from large off-premise retailers, future work could expand examinations into alcohol product availability from smaller retailers and on-premise venues.

Conclusion

While web scraping is increasingly used in the commercial sector and for consumer price index construction, its application in alcohol pricing research remains limited. This study highlights the value of web scraped datasets as a robust resource for public health policy development. Our work presents a novel, large-scale, and longitudinal approach to alcohol price data collection across a broad segment of the Australian retail market. Given the dynamic nature of alcohol pricing and the variability in product representation across retailers, this study demonstrates a viable and independently cross-checked data stream that addresses these complexities. This study contributes to the understanding of the potential use of web scraping to support the development of evidence-based alcohol policy and offers an empirical foundation for future research using this method.

CRediT authorship contribution statement

Tina Lam: Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Cynthia Huang:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Alexandra Torney:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sarah Callinan:** Writing – review & editing, Writing – original draft, Supervision, Formal analysis, Conceptualization. **Brian Vandenberg:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization. **Ting Xia:** Writing – review & editing, Formal analysis. **Colin Angus:** Writing – review & editing, Supervision. **Robin Room:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization. **Rowan Ogeil:** Writing – review & editing, Writing – original draft, Conceptualization. **Aaron Cowper:** Writing – review & editing, Supervision, Resources. **Simone Pettigrew:** Writing – review & editing, Supervision. **Bosco Rowland:** Writing – review & editing. **Danica Keric:** Writing – original draft. **Dan I Lubman:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization. **Suzanne Nielsen:** Writing – review & editing, Writing – original draft, Supervision, Resources, Funding acquisition, Conceptualization.

Declaration of competing interest

The academic authors declare that they have no known competing financial interests or personal relationships to the alcohol industry that could have appeared to influence the work reported in this paper. The academic team obtained some of the data used in this paper from commercial data aggregators ShopGrok. ShopGrok have clients from a range of retail sectors, some of which sell alcohol - these clients had no role in the study design, study conduct, analysis or data interpretation.

Acknowledgements

This study was funded by Monash University (Monash Data Futures Institute Seed Grant Program and the Advancing Women's Success Grant). The authors gratefully acknowledge Samuel Campbell and Nikhil Suresh for their technical assistance and advice in constructing the 2021 custom Monash University web scraper, and Mitchell O'Hara-Wild for his technical assistance in the construction of the 2024 Monash University validation web scrapes. We thank Dennis Petrie for his role in funding acquisition and research advice. DIL and SN are supported by an NHMRC Leadership Fellowship and SP is supported by an NHMRC Investigator Grant (APP2034602).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.drugpo.2025.105115.

References

Albis, M. L. F., Romasoc, S. O., Pelayo, S. G., Gavira, B. A. C., & Asombrado, J. P. J. (2023). Web scraping for price statistics in the Philippines. *Statistical Journal of the IAOS, 39*, 933–945.

Australian Bureau of Statistics (ABS). Consumer Price Index: Concepts, sources and methods. from https://www.abs.gov.au/statistics/detailed-methodology-information/concepts-sources-methods/consumer-price-index-concepts-sources-and-methods/2018/price-collection.

Australian Government. (2018). Final budget outcome 2017-18. In.

Bentley, A., Chipeniuk, K., & Krsinich, F. (2024). Online price data, high frequency inflation measures, and formulation of monetary policy in New Zealand.

Bertin, L., Leung, G., Bohm, M. K., LeClercq, J., Skillen, E. L., & Esser, M. B. (2024). Estimating the effects of hypothetical alcohol minimum unit pricing policies on alcohol use and deaths: A state example. *Journal of Studies on Alcohol and Drugs, 85*, 120–132.

Brauer, M., Roth, G. A., Aravkin, A. Y., Zheng, P., Abate, K. H., Abate, Y. H., … Gakidou, E. (2024). Global burden and strength of evidence for 88 risk factors in 204 countries and 811 subnational locations: A systematic analysis for the Global Burden of Disease Study 2021. *The Lancet, 403*, 2162–2203.

Brennan, A., Meng, Y., Holmes, J., Hill-McManus, D., & Meier, P. S. (2014). Potential benefits of minimum unit pricing for alcohol versus a ban on below cost selling in England 2014: Modelling study. *British Medical Journal, 349*, g5452.

Burton, R., Henn, C., Fitzgerald, N., & Sheron, N. (2024). The early impact of the UK's new alcohol taxation system on product strength and price: An exploratory comparative descriptive study. *Public Health, 232*, 61–67.

Callinan, S., Laslett, A.-M., Rekve, D., Room, R., Waleewong, O., Benegal, V., Casswell, S., Florenzano, R., Hanh, H. T. M., Hanh, V. T. M., Hettige, S., Huckle, T., Ibanga, A., Obot, I., Rao, G., Sieng Sounthone, L., Rankin, G., & Thamarangsi, T. (2016). Alcohol's harm to others: An international collaborative project. *International Journal of Alcohol and Drug Research, 5*, 8.

Callinan, S., Room, R., Livingston, M., & Jiang, H. (2015). Who purchases low-cost alcohol in Australia? *Alcohol and Alcoholism, 50*, 647–653.

Cavallo, A., & Rigobon, R. (2016). The billion prices project: Using online prices for measurement and research. *Journal of Economic Perspectives, 30*, 151–178.

Chalmers, J., Carragher, N., Davoren, S., & O'Brien, P. (2013). Real or perceived impediments to minimum pricing of alcohol in Australia: Public opinion, the industry and the law. *International Journal of Drug Policy, 24*, 517–523.

Davies, T., Lam, T., & Pettigrew, S. (2025). Alcohol pricing and discounting in Western Australia: Identifying products below hypothetical minimum unit prices. *Drug and Alcohol Review, 44*, 2013–2018.

Davies, T., O'Brien, P., Bowden, J., Petticrew, M., & Pettigrew, S. (2024). The prevalence and magnitude of price promotions in online alcohol retail outlets. *Drug and Alcohol Review, 43*, 1705–1709.

Economics in Context Initiative. (2021). *Comparative economic systems: Capitalism and socialism in the 21st century*. Global Development Policy Center, Boston University.

Foerderer, J. (2023). Should we trust web scraped data?. *Working paper*. Technical University of Munich.

Fogarty, A., & Chapman, S. (2013). What should be done about policy on alcohol pricing and promotions? Australian experts' views of policy priorities: A qualitative interview study. *BMC Public Health, 13*, 610-610.

Gill, J., Chick, J., Black, H., Rees, C., O'May, F., Rush, R., & McPake, B. A. (2015). Alcohol purchasing by ill heavy drinkers; cheap alcohol is no single commodity. *Public Health, 129*, 1571–1578.

Gruenewald, P. J., Ponicki, W. R., Holder, H. D., & Romelsjö, A. (2006). Alcohol prices, beverage quality, and the demand for alcohol: Quality substitutions and price elasticities. *Alcoholism, Clinical and Experimental Research, 30*, 96–105.

Hill-McManus, D., Brennan, A., Stockwell, T., Giesbrecht, N., Thomas, G., Zhao, J., Martin, G., & Wettlaufer, A. (2012). *Model-based appraisal of alcohol minimum pricing in Ontario and British Columbia: A Canadian adaptation of the Sheffield alcohol policy model version 2*. University of Sheffield.

Hillen, J. (2019). Web scraping for food price research. *British Food Journal, 121*, 3350–3361.

Holmes, J., Meng, Y., Meier, P. S., Brennan, A., Angus, C., Campbell-Burton, A., Guo, Y., Hill-McManus, D., & Purshouse, R. C. (2014). Effects of minimum unit pricing for alcohol on different income and socioeconomic groups: A modelling study. *Lancet, 383*, 1655–1664.

Huang, C., & Lam, T. Prites: An integrative framework for investigating and assessing web-scraped http-response datasets for research applications. Retrieved 19 November 2025 from https://arxiv.org/abs/2511.13773.
Jiang, H., Callinan, S., Livingston, M., & Room, R. (2017). Off-premise alcohol purchasing in Australia: Variations by age group, income level and annual amount purchased. *Drug and Alcohol Review, 36*, 210–219.
Jiang, H., & Livingston, M. (2015). The dynamic effects of changes in prices and affordability on alcohol consumption: An impulse response analysis. *Alcohol and Alcoholism, 50*, 631–638.
Jones, S., Barrie, L., Robinson, L., Allsop, S., & Chikritzhs, T. (2012). Point-of-sale alcohol promotions in the Perth and Sydney metropolitan areas. *Drug and Alcohol Review, 31*, 803–808.
Khder, M. (2021). Web scraping or Web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing and its Applications, 13*, 145–168.
Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*, 155–163.
Kruskal, W., & Mosteller, F. (1979). Representative sampling, II: Scientific literature, excluding statistics. *International Statistical Review, 47*, 111–127.
Lam, T., Callinan, S., Nielsen, S., Horn, F., Francia, L., & Vandenberg, B. (2023). Which alcohol products might be affected by the introduction of a minimum unit price in Western Australia? Findings from a survey of alcohol retail prices. *Drug and Alcohol Review*.
Lam, T., Vandenberg, B., Callinan, S., Horn, F., Francia, L., Campbell, S., Suresh, N., & Nielsen, S. (2021). *WA alcohol retailer price audit: Research report on the implications of a minimum unit pricing policy*. Perth: Cancer Council.
Lusk, J. L., & Brooks, K. (2011). Who participates in household scanning panels? *American Journal of Agricultural Economics, 93*, 226–240.
Monash Partners. Research collaboration agreement Monash Partners template. Retrieved 26 August 2025 from https://monashpartners.org.au/wp-content/uploads/2024/02/Research-Collaboration-Agreement-Monash-Partners-Template-19-02-2024.docx.
Morrison, C., Smith, K., Gruenewald, P. J., Ponicki, W. R., Lee, J. P., & Cameron, P. (2016). Relating off-premises alcohol outlet density to intentional and unintentional injuries. *Addiction, 111*, 56–64.
Murti, L. (2024). A supermarket trip may soon look different, thanks to electronic shelf labels. *2024*.
O'Donnell, A., Anderson, P., Jané-Llopis, E., Manthey, J., Kaner, E., & Rehm, J. (2019). Immediate impact of minimum unit pricing on alcohol purchases in Scotland: Controlled interrupted time series analysis for 2015-18. *British Medical Journal, 366*, l5274.
Office of Impact Analysis. Developing the evidence base for decision making. from https://oia.pmc.gov.au/.
Parkes-Hupton, H. Major Aussie bottleshop chains to vanish completely. Retrieved March 26 2025 from https://www.heraldsun.com.au/business/companies/major-aussie-bottleshop-chains-to-vanish-completely/news-story/064d21c88999fa52d24b5bcc11974645.
Profaca, C. What are the benefits of electronic shelf labelling systems for liquor retail? Retrieved 12 November 2024 from https://www.drinkstrade.com.au/news/what-are-the-benefits-of-electronic-shelf-labelling-systems-for-liquor-retail.
Rankin, G., & Livingston, M. (2016). *Understanding alcohol sales data in Australia*. Melbourne, Australia: Foundation for Alcohol Research and Education.
Ravandi, B., Ispirova, G., Sebek, M., Mehler, P., Barabasi, A. L., & Menichetti, G. (2025). Prevalence of processed foods in major US grocery stores. *Nature Food*.
Reeves, M. (2021). Liquor retailing in Australia (November 2021). In.
Reeves, M. (2025) Online beer, wine and liquor sales in Australia - market research report (2014-2029). from https://www.ibisworld.com/australia/industry/online-beer-wine-and-liquor-sales/4087/.
Rehm, J., Kilian, C., Rovira, P., Shield, K. D., & Manthey, J. (2021). The elusiveness of representativeness in general population surveys for alcohol. *Drug and Alcohol Review, 40*, 161–165.
Rennie, S., Buchbinder, M., Juengst, E., Brinkley-Rubinstein, L., Blue, C., & Rosen, D. L. (2020). Scraping the web for public health gains: Ethical considerations from a 'Big data' Research project on HIV and incarceration. *Public Health Ethics, 13*, 111–121.
Roy Morgan. (2019) Independent retailers gain in packaged alcohol market. from https://www.roymorgan.com/findings/independent-retailers-gain-in-packaged-alcohol-market.
Seaton, J. S., & Waterson, M. (2013). Identifying and characterising price leadership in British supermarkets. *International Journal of Industrial Organization, 31*, 392–403.
Secombe, P., Campbell, L., Brown, A., Bailey, M., & Pilcher, D. (2021). Hazardous and harmful alcohol use in the Northern Territory, Australia: The impact of alcohol policy on critical care admissions using an extended sampling period. *Addiction, 116*, 2653–2662.
Sharma, A., Sinha, K., & Vandenberg, B. (2017). Pricing as a means of controlling alcohol consumption. *British Medical Bulletin, 123*, 149–158.
Srivastava, P., Yang, O., & Zhao, X. (2022). Equal tax for equal alcohol? Beverage types and antisocial and unlawful behaviours. *Economic Record, 98*, 354–372.
Stockwell, T., Churchill, S., Sherk, A., Sorge, J., & Gruenewald, P. (2020). How many alcohol-attributable deaths and hospital admissions could be prevented by alternative pricing and taxation policies? Modelling impacts on alcohol consumption, revenues and related harms in Canada. *Health Promotion and Chronic Disease Prevention in Canada, 40*, 153–164.
Stockwell, T., Giesbrecht, N., Vallance, K., & Wettlaufer, A. (2021). Government options to reduce the impact of alcohol on Human health: Obstacles to effective policy implementation. *Nutrients, 13*.
Whetton, S., Tait, R. J., Gilmore, W., Dey, T., Agramunt, S., Halim, S. A., McEntee, A., Mukhtar, A., Roche, A., Allsop, S., & Chikritzhs, T. (2021). *Examining the social and economic costs of alcohol use in Australia*. Perth, WA: NDRI, Curtin University.
World Health Organization. (2019). *The safer technical package: Five areas of intervention at national and subnational levels*. Geneva: World Health Organization.
Roy Morgan Research (2017). The Australian alcohol retail market in review. Retrieved 2021 from https://www.roymorgan.com/findings/the-australian-alcohol-retail-market-in-review.
World Health Organization. (2024). Global status report on alcohol and health and treatment of substance use disorders. In.
Zhao, J., Stockwell, T., Martin, G., Macdonald, S., Vallance, K., Treno, A., Ponicki, W. R., Tu, A., & Buxton, J. (2013). The relationship between minimum alcohol prices, outlet densities and alcohol-attributable deaths in British Columbia, 2002-09. *Addiction, 108*, 1059–1069.
Zorbas, C., Lee, A., Peeters, A., Lewis, M., Landrigan, T., & Backholer, K. (2021). Streamlined data-gathering techniques to estimate the price and affordability of healthy and unhealthy diets under different pricing scenarios. *Public Health Nutrition, 24*, 1–11.