



Worldpanel By Numerator Data Processing Report

Colin Angus on behalf of the Sheffield Addictions Research Group

August 2026

DOI: <https://doi.org/10.15131/shef.data.33009752>

Introduction

This report details the data cleaning/processing pipeline developed by the Sheffield Addictions Research Group (SARG) with data provided by Worldpanel by Numerator (formerly Kantar). It describes the steps taken in producing the final analytical dataset used in all analyses. All analyses and conclusions in this report are made by SARG independently of Worldpanel by Numerator.

The data

The raw data is provided by market research company Worldpanel by Numerator's Take Home Purchase panel. This is a continuous, rolling panel of approximately 30,000 households in Great Britain. Households in the panel record all food and drink purchases that they bring into their home. SARG has purchased data relating to purchases of all alcohol and no/lo drinks (where no/lo drinks are beers, ciders, wines, spirits and alcopops with an alcohol content of no more than 1.2% ABV) for the period from 01 January 2018 to 31 December 2025. The data is at the transaction level, with information about the product, including brand, pack size and type and ABV, the volume purchased and the total price paid. Each individual purchase is a single row in the data, so if a household bought (and brought home) 4 cans of brand x lager, 12 bottles of brand y lager and 2 bottles of brand z wine at the same time, this would appear as 3 separate rows in the data. We refer to each row as a 'transaction'. The data also includes information about the purchasing households - geographical region, household composition and sociodemographic details about the household's primary shopper.

Panellists do not declare to Worldpanel by Numerator the price they paid for each individual item. Rather, the data providers have visibility through receipts from a proportion of shopping trips and then apply a process to impute prices across the remainder of the dataset. This process uses receipt data where available and an algorithm for the remaining records, taking into account factors such as the number of available data records. Smaller numbers of data records may result in greater variation in price estimates, as there are fewer observations available to match against.

In total, the data held by SARG covers 6.1million individual purchase transactions from approximately 40000 separate households.

The data cleaning process

In order to prepare the data for analysis, SARG undertake a series of data checks to identify potential issues in the raw data. Where issues are identified, these are corrected where there is a clear rationale for editing some aspect of the raw transaction, or flagged where no obvious edit can be identified. Flagged transactions can then be excluded from analyses where appropriate, although this may not always be the case - for example products whose price has been flagged may still be included in analyses of sales volumes.

Here we step through the various data checks and edits applied. Numerator's disclosure control means that we are unable to identify the specific brands, products or transactions that have been identified at each step, but we provide general details to illustrate the process.

Data checks

Transactions with missing ABV data

In total 0.015% of all transactions have missing ABV data. The vast majority of these transactions come from a small number of products. We filled in missing ABVs using the following hierarchy: 1) Where all other transactions had a single non-missing ABV value for the same product, we used that 2) Where other transactions had multiple non-missing ABV values for the same product, we compared transactions, looking for trends over time, supplemented with online searches for the product, to identify the most plausible ABV 3) Where no other transactions were available in the data, we searched online for the most plausible ABV 4) Where we could not identify a plausible ABV from any of these sources

(e.g. where the product was a generic own-brand) we used the median ABV in the data for similar products (e.g. products from the same category at a similar price point).

This process allowed us to establish a most plausible ABV for all missing values.

Zero value transactions

In total 7000 transactions in the data have a sales value of £0, 0.115% of all transactions. The vast majority of these have a promotional offer recorded against them e.g. 'buy 3 get one free'. This reflects that fact that for these offers, the raw data records this as several products at full price and one 'free'. We instead identify all of the products included in the promotion (by matching transactions from the same household on the same day with the same promotional flag), sum their purchase value and divide it equally between the included transactions. For example, if a household had bought 3 bottles of wine for £8 each and received a 4th bottle free, we would treat this as 4 bottles at £6 each ($8 \times 3 / 4$). This approach removes the vast majority of zero-value transactions, but not all, with 0.002% of all transactions remaining with zero values in the final cleaned data. These reflect transactions which either do not have one of these promotional offers recorded against them, or where we cannot identify a corresponding set of transactions to combine it with (e.g. a household that recorded two bottles of wine purchased on the same date, one for £5 and one for £0 with a 'buy 4 get one free' promotion). These transactions are flagged for possible removal in analyses where appropriate.

Alcohol transactions recorded below the cost of duty + VAT

For sales of alcoholic products above 1.2% ABV, the UK Government introduced legislation in 2014 that made it illegal for the sales price to be lower than the cost of the alcohol duty payable on the product, plus the VAT on that duty. The intention of this legislation was to prevent retailers from 'loss-leading' on alcohol - selling products at a loss to drive traffic to their stores - and it is sometimes referred to as a 'ban on below-cost selling' as a result. This legislation only applies to England & Wales, however we consider it very unlikely that there are significant sales below this threshold in Scotland, so we apply this check to the full dataset.

In total 26300 transactions in the data for alcoholic products have a sales price above £0, but below the applicable level of duty plus VAT on that duty, 0.443% of all alcohol purchase transactions. These low-value transactions come from a much more diverse set of products than the missing ABVs. Our approach here was to manually inspect the transactions for every product with at least 100 transactions below this threshold.

These inspections identified 4 common themes:

- 1) Products, particularly high-strength lagers, where the sales price of some products listed as 4-packs strongly suggests that the actual purchase was for a single can (e.g. where the majority of sales for that product are £6-8, but some transactions appear at £2). For these products we identified the most plausible threshold below which we believed that sales most likely represent single cans and manually recoded all sales below this threshold to be single cans.
- 2) Products, particularly wines, where the product is listed as a standard 750ml bottle, but the sales price strongly suggests that the sale was a smaller 250ml or 187.5ml bottle. For these products, where we could confirm that a smaller-sized bottle did exist, we identified the most plausible threshold below which we believed that sales most likely represent the smaller bottle size and manually recoded all sales below this threshold to the smaller bottle size.
- 3) Products where some transactions appear to have been allocated an incorrect ABV (e.g. a fortified wine which was commonly listed at 15%, but some transactions at a similar price point and pack size were listed at 30%). For these we manually recoded the ABV of all transactions to the most plausible ABV value given the sales price recorded.
- 4) Products where we could not readily identify a single most plausible cause for the low sales price, or where there was no single edit that would bring the sales price above the 'below cost' threshold that was more plausible than any other. Since changing any of the product price, volume or ABV could alter either the sales price or the level of duty payable. These products were left in the data unchanged, but were flagged as low priced.

In general we employed a cautious approach, only editing the raw data where we believed it was likely that the edit represented a highly plausible change. After applying these edits, 0.272% of transactions remained in the data with non-zero, but low sales prices, all of which were flagged.

Alcohol transactions below the Minimum Unit Price threshold

Scotland has had an MUP, preventing the sale of alcohol for less than 50p per unit since 1st May 2018. This was increased to 65p on 30th September 2024. Wales has had an MUP of 50p per unit since 2nd March 2020. If we exclude any products identified as having implausibly low prices in relation to their tax value (as detailed above), then a total of 20200 transactions

in the data in a country with MUP have been recorded at a sales price below the relevant MUP threshold. This corresponds to 2.87% of all transactions in countries under MUP.

There is some variation between the various phases of MUP implementation. Of these transactions, 8400 were in Scotland under the original 50p MUP threshold, 1600 were since the threshold was raised to 65p and 10100 were in Wales since the introduction of their 50p MUP. These correspond to 1.98%, 2.78% and 4.6% of all transactions respectively in each country/period. We note that it is possible that some of these below-MUP transactions are the result of cross-border shopping - households in Scotland or Wales that have purchased alcohol in England, where MUP does not apply. However, while we do not have robust data on the proportion of shopping trips that are cross-border, the very small proportion of the Scottish population in particular who live close to the English border suggests that this is unlikely to explain a substantial proportion of these below-MUP transactions.

As with the transactions identified below the cost of duty plus VAT on the duty, we manually inspected products with at least 100 transactions below the relevant MUP level. We employed the same approach as for transactions below the tax value, only making edits where we could identify a highly plausible change. There were far fewer of these than for the below tax transactions, reflecting the fact that the majority of those below the MUP threshold were only narrowly so. Following any edits (including the edits described in previous sections), 2.86% of transactions taking place under an MUP policy remained below the MUP level. Note that this figure reflects both a reduction in the number of below-MUP transactions from these edits, but also potentially an increase due to the edits described and applied elsewhere (e.g. those to address below-tax transactions).

For these remaining below-MUP transactions, there are two possible approaches. Firstly, we create a version of the data in which we increase the recorded sales price to exactly the level required to bring the price per unit of alcohol up to the MUP threshold. Secondly, we create an alternative version of the original data in which we flag the transactions and these can either be removed, or retained, for any analysis.

Transactions with unusually high prices

As well as checking for low-priced transactions, we also inspect the data for transactions where the recorded price exceeds the levels that we would expect to see in alcohol sales data. We perform these checks on the basis of the sales price per litre of product, rather than per unit of alcohol as similar plausibility thresholds can then be applied to both alcoholic and no/lo products. For each alcohol category (beer, cider, RTDs (Ready-To-Drinks - pre-mixed spirit- or wine-based drinks including alcopops), spirits and wine) inspection of the price

distribution suggested that prices generally increase sharply above the 99.9th percentile of prices. We therefore manually inspected all transactions for products above this level that accounted for at least 10 transactions. This identified a small number of plausible edits to the data - for example some bottles of spirits are recorded as being 35ml but are priced at a level more in line with a 350ml bottle - which we applied using the same cautious approach outlined above. We also established maximum plausible price thresholds of £12 per litre for beer and cider, £65 per litre for RTDs and £100 per litre for wine. A small number of very high spirits prices appear to be genuine and we therefore could not establish a maximum price thresholds for spirits.

In total 0.024% of all transactions were above these thresholds in the original data. After applying any identified edits, this was reduced to 0.014. These remaining high-priced transactions were flagged.

Transactions with unusually low pack sizes

The final data check is to manually inspect any transactions that appear below the smallest container sizes that we would expect to find in the UK market. This included beers below 250ml, cider below 200ml (as our cider category also includes perry that sometimes comes in 200ml containers), RTDs below 100ml, spirits below 50ml and wine below 100ml. Overall, 0.001% of all transactions fell below these thresholds. As with other checks, plausible edits were made to the pack sizes of these products, taking a cautious approach. These edits did not leave any remaining transactions below these volume thresholds.

Headline impacts of edits applied to the data

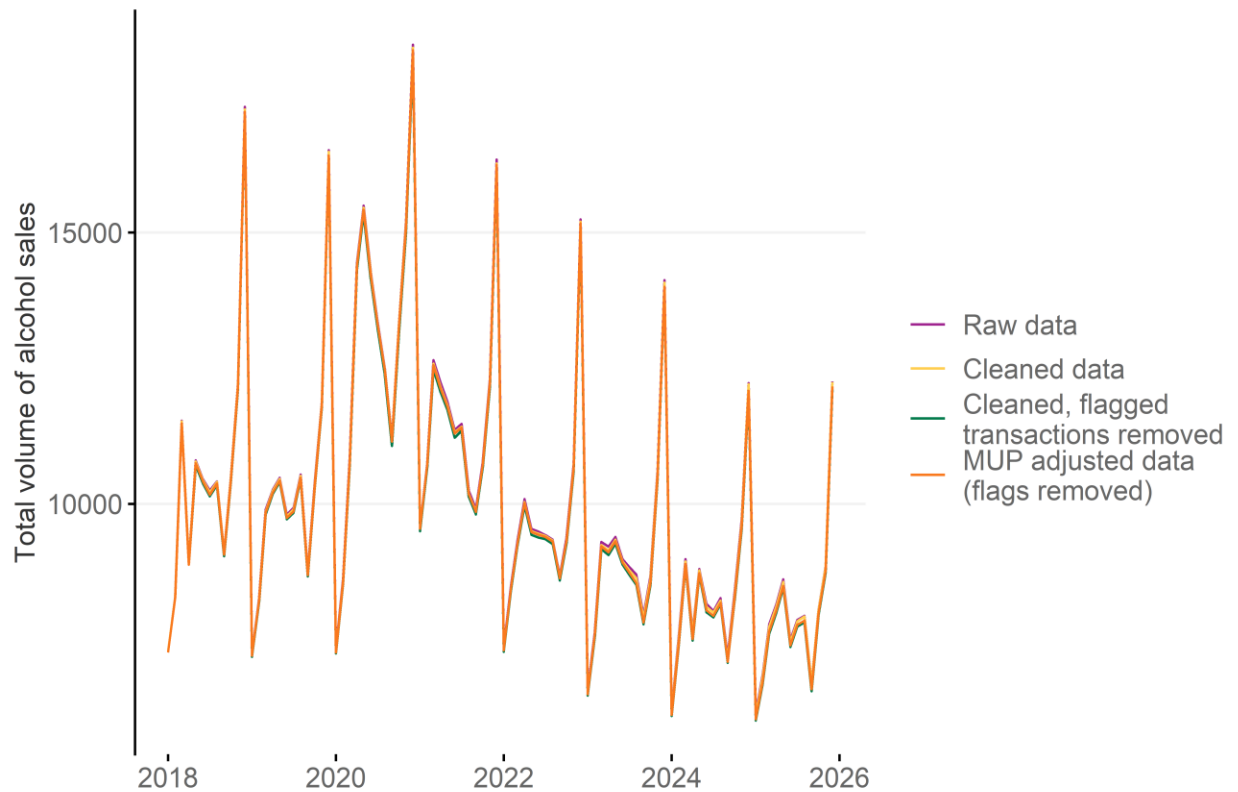
These edits leave us with four alternative versions of the Worldpanel by Numerator data: 1) The original, raw, data 2) The cleaned data with flagged transactions left in 3) The cleaned data with flagged transactions removed 4) The cleaned data with below-MUP priced transactions increased to the MUP threshold and other flagged transactions removed

Here we briefly compared the differences between each version of the data in relation to common key metrics, to assess the extent to which the edits we have applied have impacted on the data. Note that versions 3 and 4 of the data are identical for no/lo products, as MUP does not apply to these.

Volume of alcohol purchased

There is very little difference in the alternative versions of the data between the total volume of alcohol recorded as being sold in the data. Removing the flagged data has the largest impact, bringing the total volume down slightly.

Litres of alcohol recorded

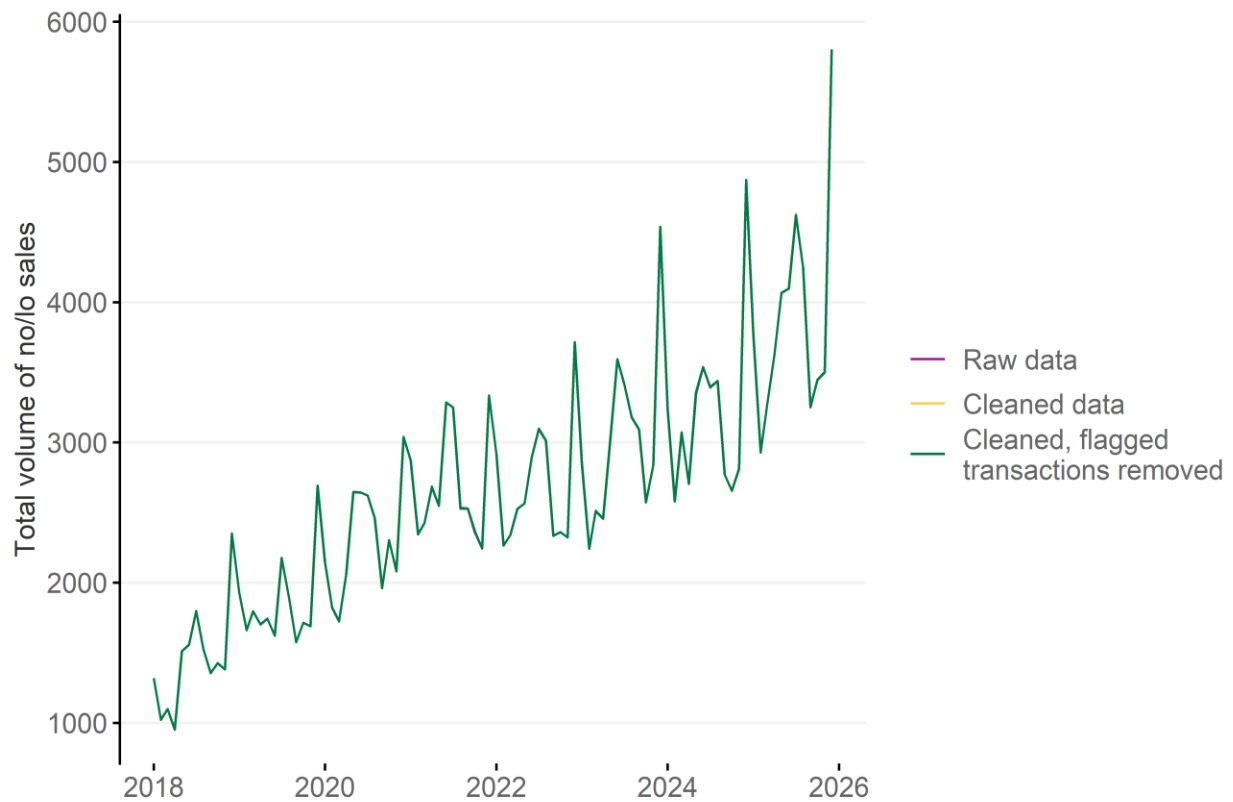


Analysis using Worldpanel by Numerator, GB Take Home Panel, Total Alcohol, 418 w/e 31 December 2025

Volume of no/lo products purchased

There is almost no difference in the alternative versions of the data between the total volume of alcohol recorded as being sold in the data.

Litres of no/lo sales recorded

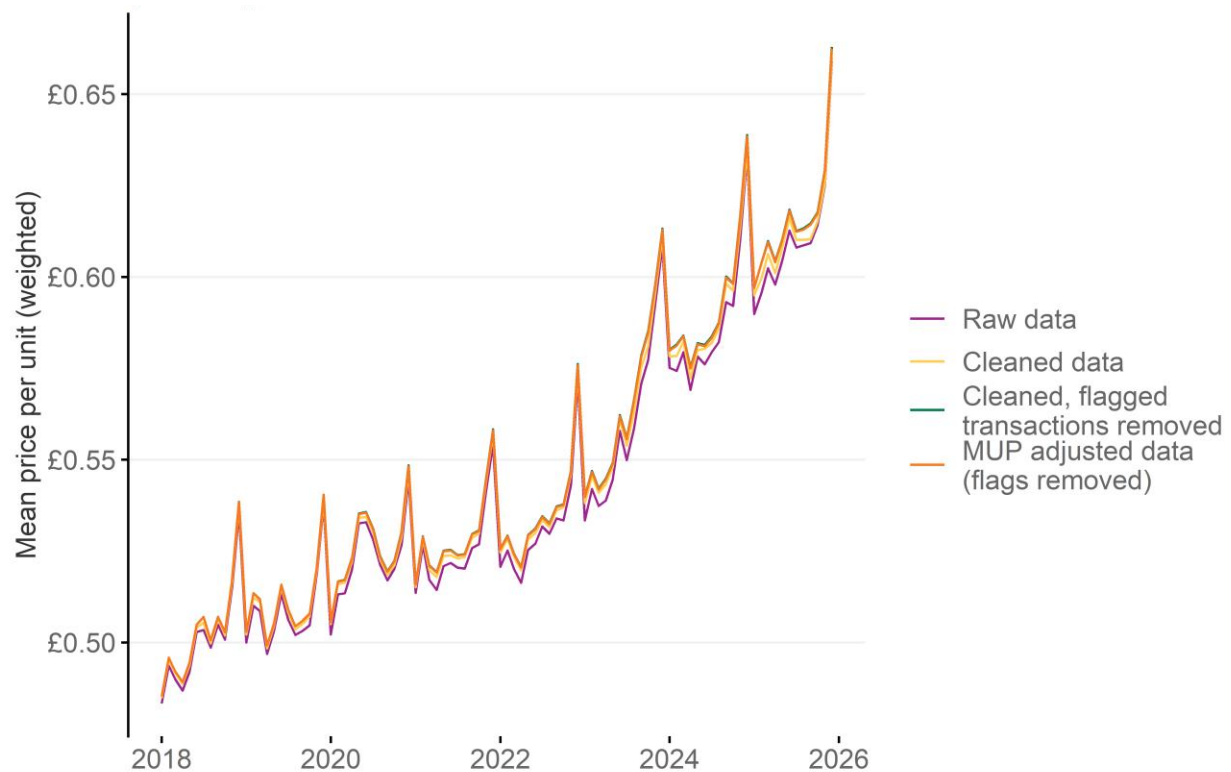


Analysis using Worldpanel by Numerator, GB Take Home Panel, Total No/Lo, 418 w/e 31 December 2025

Mean price per unit of alcohol

Plotting the mean price paid per unit of alcohol (weighted by volume of alcohol) each month shows small differences, with the mean price in the cleaned data being marginally higher than the raw data, and the cleaned data with flagged transactions removed being higher still. This is unsurprising, as the majority of edits applied to the data are applied to increase the prices of lower-priced transactions.

Mean price per unit

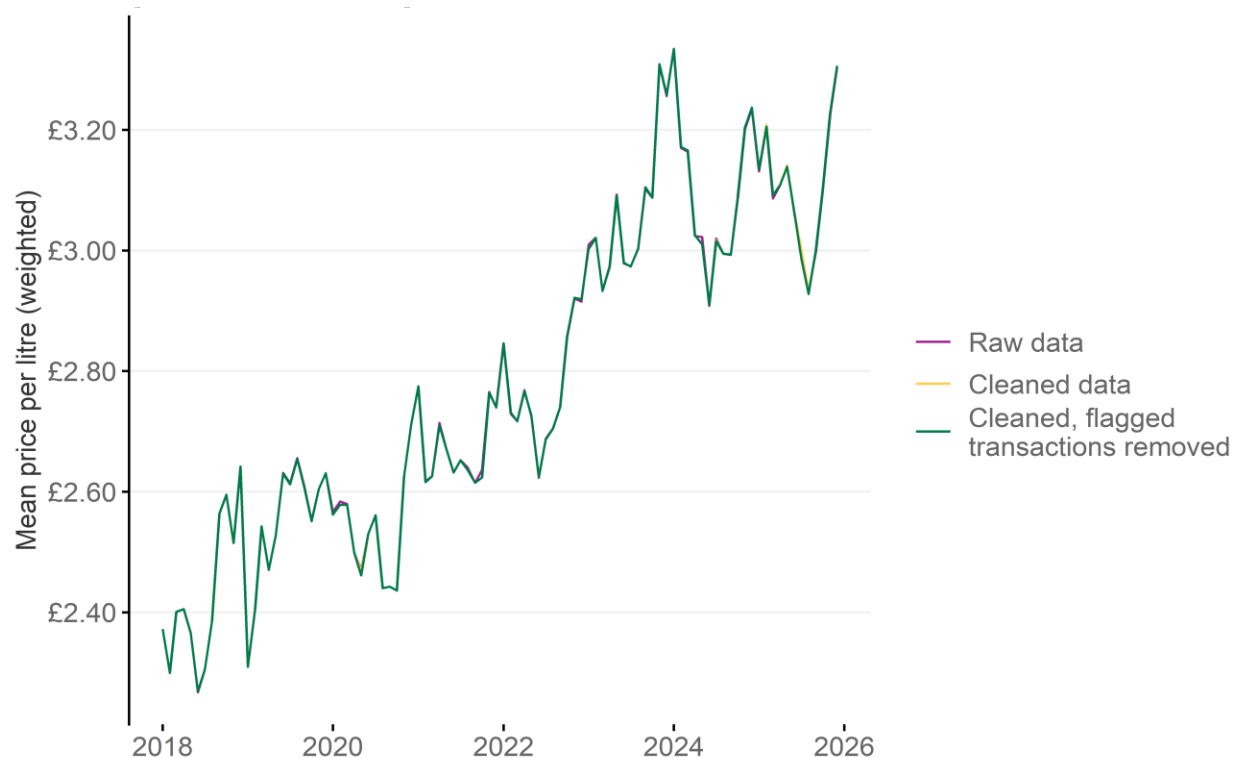


Analysis using Worldpanel by Numerator, GB Take Home Panel, Total Alcohol, 418 w/e 31 December 2025

Mean price per litre of no/lo alcohol products

Plotting the mean price paid per litre of no/lo products (weighted by the volume of product) sold each month shows little difference, again, between the versions of the data.

Mean price of no/lo products



Analysis using Worldpanel by Numerator, GB Take Home Panel, Total No/Lo, 418 w/e 31 December 2025

Conclusion

This report outlines the data cleaning processes applied by the Sheffield Addictions Research Group to Worldpanel by Numerator data. We describe a series of checks and data validation steps and outline the approach to identifying and applying subsequent edits to the raw data based on these checks. Collectively these edits have only a small impact on the volume of alcohol sold and the average price paid for alcohol products and almost no impact on sales volumes or prices for no/lo products.