P1067 SD 1 Appendix 3. Summary of study characteristics

Australian Studies

Study (authors/ date)	Date data collected	Study type	Population	Design	HSR stimuli shown	Quality Assessment
Acton et al. (2023)^	2018, 2019, 2020	Quantitative	N = 64,032 adults from five countries including 12,418 Australians. Nationally representative of age, gender and education; shopping status not reported	Annual repeated online cross-sectional surveys exploring self-reported awareness, use and understanding of front-of-pack and nutrition fact labels	HSR only (not on product)	High Rated high on data collection tools and analysis and moderately on sample representativeness. Rated poorly on recruitment procedures. No major concerns with overall execution, though some information is available only through external sources rather than in the article itself.
Bhawra et al. (2022a)^[1]	2018	Quantitative	N = 21,586 adults from five countries including 3,901 Australians. Quotas for age and sex based on country statistics; 71.6% primary shopper, 21.5% shared shopping, 6.9% not primary shopper			High Rated high on data collection tools and analysis and moderately on sample representativeness. Rated poorly on recruitment procedures. No major concerns with overall execution, though some information is available only through external sources rather than in the article itself.
Bhawra et al. (2022b)^[2]	2018	Quantitative	N = 21,586 adults from five countries including 3,901 Australians.			High

[1] Subset of population used in Acton 2023, i.e., population from 2018.

[2] Same population as Bhawra 2022a.

Study (authors/ date)	Date data collected	Study type	Population	Design	HSR stimuli shown	Quality Assessment
			Quotas for age and sex based on country statistics 51.3% female 41.6% low education 71.6% primary food shopper			Rated high on data collection tools and analysis and moderately on sample representativeness. Rated poorly on recruitment procedures. No major concerns with overall execution, though some information is available only through external sources rather than in the article itself.
Jáuregui et al. (2021)^[3]	2018	Quantitative	N = 22,140 adults from five countries including 3,964 Australians. Participants were recruited through the Nielsen Consumer Insights Global Panel and their partners' panels using both probability and non-probability sampling. Stratified by age and sex proportional to the general population. Of participants assigned to the HSR condition, 72.1% (99% CI: 69.8, 74.3) shopped for their household.	Six-arm RCT undertaken via online survey. Randomly assigned to view on screen one of six images of a sweetened fruit drink with differing FOP labelling.	HSR only (on product)	High Rated high on data analysis approach. Rated moderately on sample representativeness, measurement tools, and procedural detail. No major concerns with overall execution, though some information is available only through external sources rather than in the article itself.

[3] Same study population as Acton 2023 and Bhawra 2022a and 2022b.

Study (authors/ date)	Date data collected	Study type	Population	Design	HSR stimuli shown	Quality Assessment
Anderson & O'Connor (2018)	Not stated	Quantitative	N = 249 83.1% female 53.8% high school highest education completed 69.5% 17-24 yo 44.6% with income <$10,499 Convenience sample from Queensland. Participants held at least some responsibility for grocery shopping.	Experimental online cross-sectional survey - forced choice of breakfast cereals in product comparison scenarios IV: Presence or absence of the HSR - 3 comparative contexts (no label vs. no label, HSR vs. no label, and HSR vs. HSR) AND Nutritional profile of the product (low vs high)	HSR only (on product)	Medium Rated high on study design and moderately on data analysis. Rated poorly on sample representativeness, recruitment procedures, and data collection tools. Some concerns with overall execution.
Billich et al. (2018)	2017	Quantitative	N = 994 18 to 35 years of age, could read and understand English, had access to a computer and internet connection, and were located in Victoria, New South Wales, South Australia or Queensland.	RCT Participants assigned to one of five study conditions. Labels were presented to the participants in colour and displayed as a large image above the pictures of drinks. Then asked which drink they would choose from the selection.	HSR only (next to product)	High Rated moderately on data analysis, collection tools and recruitment. No major concerns with overall execution.
Brown et al. (2017)	Not stated	Quantitative	N = 117 18 to 30 years of age recruited from the University of Newcastle Callaghan campus (Nutrition and Dietetics students excluded)	RCT Adults randomly assigned to a kJ/100 g food label or a HSR label or no nutrition information and asked to serve themselves adequate portions of food.	Not stated	Medium Rated high on use of validated tools and appropriate statistical analysis. Rated poorly on recruitment reporting and detail on measurement tools. No major concerns with overall execution.

Study (authors/ date)	Date data collected	Study type	Population	Design	HSR stimuli shown	Quality Assessment
Cooper et al. (2020)	Nov 2014 - Feb 2015	Quantitative	N = 1,024 adults with quota sampling for age, gender, location. Required to be primary grocery shopper for household.	Contingent valuation (a type of discrete choice experiment) to assess how much consumers would be willing to pay for the HSR to be present on food packages.	HSR with tail (not on product)[4]	High Rated high on description of the research setting and use of appropriate, validated methods. Rated poorly on recruitment reporting and detail on measurement tools. Some minor concerns with overall execution.
FOP-ICE (Front-Of-Pack International Comparative Experimental) study						
Egnell et al. (2018)†	2018	Quantitative	N = 12,015 from 12 countries including 1,000 Australians. Quota sampling accounting for age (one-third of recruited participants in each of the following age categories: 18–30 years, 31–50 years, over 51 years), sex (50% women), and socioeconomic status (one-third of recruited participants in each of the following household income levels: low, medium, and high). 71.9% of Australians were responsible for grocery shopping and	RCT Online. Participants randomised to one of 5 FOPL conditions and asked to rank products in terms of nutritional quality.	HSR with tail (on product)	High Rated high on large sample size, cross-country study, and clarity of procedures and stimuli. Rated poorly on recruitment details, tool validation, and lack of subgroup analysis to assess equity of effects. No major concerns with overall execution.

[4] This graphic was not the official HSR logo as was provided at a time the HSR was in development.

Study (authors/ date)	Date data collected	Study type	Population	Design	HSR stimuli shown	Quality Assessment
			20.7% shared shopping equally.			
Pettigrew et al. (2023)†	2018	Quantitative	N = 1,000 Australian adults with quotas for age, gender, income and location.	RCT Respondents were shown three products of varying nutritional quality for each of the three product categories of breakfast cereals, cakes and pizzas and asked to rank them by nutritional quality.	HSR with tail (on product)	Medium Some concerns with overall execution.
Talati et al. (2019b)†	2018	Quantitative	N = 12,015 from 12 countries including 987 Australians. Quotas were applied so that the sample was evenly split according to gender, age (within the following brackets: 18–30 years, 31–50 years, >50 years) and income level (low, medium and high).	RCT Respondents were shown foods of varying nutritional quality (with no label on package) and selected which they would be most likely to purchase. The same choice sets were then shown again with 1 of 5 randomly allocated labels on package (HSR, Multiple Traffic Lights, Nutri-Score, Reference Intakes, or Warning Label).	HSR with tail (on product)	High Rated high on a large international sample, clear study design, strong procedural clarity. Rated moderately on sample representativeness, recruitment detail, tool validation, and subgroup analysis. No major concerns with overall execution.
National Heart Foundation (2017)	Three waves: Sep 2015, Feb 2016, Jul 2016	Qualitative/ Quantitative	N= 6,044 Three waves: Sept 2015 (n=2,036), Feb 2016 (n=2,005), July 2016 (n=2,003)	Online cross-sectional survey investigating consumer awareness, use, understanding and trust in HSR	All forms (not on product)	Medium Rated high on setting description and survey tool detail. Rated poorly on sampling, recruitment, analysis, and tool validity.

Study (authors/ date)	Date data collected	Study type	Population	Design	HSR stimuli shown	Quality Assessment
			Adult participants required to be main or shared grocery buyer			Concerns with overall execution.
National Heart Foundation (2019)	2015 to 2018	Qualitative/ Quantitative	Year 2: N = 4,041 Sep 2015 (n = 2,036); Feb 2016 (n = 2,005) Year 3: N = 7,010 Jul 2016 (n = 2,003)[5]; Dec 2016–Mar 2017 (n = 2,507); Apr–July 2017 (n = 2, 501) Year 4: N = 7,481 Aug–Nov 2017 (n = 2,531); Dec 2017–Mar 2018 (n= 2,514); Apr–Jul 2018 (n= 2,436)	Online cross-sectional survey investigating consumer awareness, use, understanding and trust in HSR Focus groups also conducted	All forms (not on product)	Medium Rated high on setting description and survey tool detail. Rated poorly on sampling, recruitment, analysis, and tool validity. Concerns with overall execution.
Neal et al. (2017)	2014 to 2016	Quantitative	N = 1,578 Adults with mean age 38 years, 84% female, had completed tertiary education in 77.5% of cases, and mostly came from households with a combined average income of >A$50,000.	RCT Individuals were randomly assigned to access one of four different formats of nutrition labelling—HSR, multiple traffic light labels, daily intake guides, warnings or control (the nutrition information panel) and were encouraged to use the smartphone application every time that they	HSR with tail (shown when scanned product)	High Rated high on statistical analysis. Rated moderately on sample representativeness, recruitment reporting, and procedural clarity. Rated poorly on measurement tool development and validation. No major concerns with overall execution.

[5] Results from Sep 2015, Feb 2016 and Jul 2016 were also presented in the National Heart Foundation 2017 report.

Study (authors/ date)	Date data collected	Study type	Population	Design	HSR stimuli shown	Quality Assessment
				purchased packaged food for 4 weeks		
Pelly et al. (2020)	Not stated	Qualitative	N = 15 A convenience sample of adults 18 years and older, reported to do at least half of the shopping for themselves or their household. Those with nutrition education were excluded.	4 x semi structured, in-person focus groups	HSR with tail (not on products and on products)	Medium Rated moderately on procedural clarity and development of the question guide. Rated poorly on sample representativeness, recruitment reporting, and depth of analysis. No major concerns with overall execution, some concerns with data analysis, which are not atypical for qualitative studies.
Pettigrew et al. (2020a)	Not stated	Quantitative	N = 1,033. Nationally representative according to age, gender and state.	RCT Participants randomly allocated to one of four conditions: black and white (B&W) full HSR, colour full HSR, B&W star rating only, and colour star rating only. Each respondent viewed four packs of breakfast cereal of varying levels of healthiness as indicated by the displayed HSRs. Respondents nominated the product option they would prefer to buy and then the one they understood to be healthiest.	HSR only (on product) and HSR with tail (on product)	Low Rated poorly on recruitment detail, tool validation, analytical adjustments, and reporting of ethics approval.

Study (authors/ date)	Date data collected	Study type	Population	Design	HSR stimuli shown	Quality Assessment
Consumer Evaluations of Pre-Packed Foods that Systematically Vary by Front-of-Pack Label, Health Claim, Price and Healthiness[6]						
Pettigrew et al. (2017)	Not stated	Qualitative/ Quantitative	N = 2,058 including 1,558 adults and 500 children aged 10 years and older Half the respondents (50%) were male and half were of low socioeconomic status (49%) as indicated by residential postcode.	Discrete choice experiment online 1 practice choice task using a nonrelated food product (muesli bars) and 8 experimental choice tasks. Participants viewed images of food packs manipulated across five attributes: front-of-pack label, health claim, price, healthiness, and food type.	HSR with tail (on product)	Medium Rated moderately on sample variation, procedural clarity, and tool development. Rated poorly on recruitment reporting, statistical transparency, and depth of open-ended analysis. Some concerns with overall execution.
Talati et al. (2016c)	2016	Quantitative	N = 1,984 (subset of Talati et al. 2017a) Quotas for age, gender and SES. Children (23%) and low SES (48%) were a focus for sampling.	Discrete choice experiment online 1 practice choice task using a nonrelated food product (muesli bars) and 8 experimental choice tasks. Participants viewed images of food packs manipulated across five attributes: front-of-pack label, health claim, price, healthiness, and food type.		Medium Rated high on statistical analysis methods. Rated moderately on sample variation and procedural reporting. Rated poorly on recruitment and measurement tool development. Some concerns with overall execution.
Talati et al. (2017a)	2016	Quantitative	N = 2,058 50% were female, 25% were children (10–17	Discrete choice experiment online		Medium

[6] The following 6 studies draw on this same online survey; Associated protocol: Talati et al. 2017. Protocol for a randomized trial assessing consumer evaluations of prepackaged foods that systematically vary by nutrition information and product attributes. BMC Nutr 3, 3.
Trial registration (retrospective): https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=370675&isReview=true

Study (authors/ date)	Date data collected	Study type	Population	Design	HSR stimuli shown	Quality Assessment
			years of age) and 49% were from low SES neighbourhoods.	1 practice choice task using a nonrelated food product (muesli bars) and 8 experimental choice tasks. Participants viewed images of food packs manipulated across five attributes: front-of-pack label, health claim, price, healthiness, and food type.		Rated moderately on sample variation and procedural detail. Rated poorly on recruitment transparency, tool validation, and analytic clarity. Some concerns with overall execution
Talati et al. (2017b)	2016	Quantitative	N = 2,069 Adults and children. Adequate representation by sex and age. Respondents residing in areas of low socioeconomic position (Socio-Economic Indexes for Areas deciles 1–4) were deliberately oversampled to reflect their higher risk of diet-related diseases.	Discrete choice experiment online 1 practice choice task using a nonrelated food product (muesli bars) and 8 experimental choice tasks. Participants viewed images of food packs manipulated across five attributes: front-of-pack label, health claim, price, healthiness, and food type.		Medium Rated moderately on sample variation, survey procedures, and alignment of analytic approach with the study design. Rated poorly on recruitment, measurement tool validation, and handling of key analytic variables. Some concerns with overall execution.
Talati et al. (2018a)	2016	Quantitative	N = 2,069 Adults and children. Adequate representation by sex and age. Respondents residing in areas of low socioeconomic position (Socio-Economic Indexes for Areas deciles 1–4) were	Discrete choice experiment online 1 practice choice task using a nonrelated food product (muesli bars) and 8 experimental choice tasks. Participants viewed images of food packs manipulated across five attributes: front-of-pack		Medium Rated moderately on sample variation, recruitment, and clarity of survey structure and stimuli. Rated poorly on measurement tool validation and analytical depth, with limited attention to key manipulated variables.

Study (authors/ date)	Date data collected	Study type	Population	Design	HSR stimuli shown	Quality Assessment
			deliberately oversampled to reflect their higher risk of diet-related diseases.	label, health claim, price, healthiness, and food type.		Some concerns with overall execution.
Talati et al. (2018b)	2016	Quantitative	N = 1,505 Respondents had an equal gender split, a skew towards those from neighbourhoods of low socio-economic status (48%).	Online cross-sectional survey Respondents viewed mock packages of unhealthy variations of pizzas, cookies, yoghurts and cornflakes featuring the DIG, MTL, HSR or no FOP label, and indicated the portion size they believed should be eaten of each food on a single occasion.		Medium Rated moderately on sample variation, clarity of survey tasks and measures, and analytical approach. Rated poorly on recruitment reporting, with limited detail on procedures and participant flow. Some concerns with overall execution.
Pollinate Research (2015)	2014 and 2015	Quantitative	N = 1,000 Nationally representative sample of main/joint grocery buyers aged 18 years and older, with a population representative of Aboriginal and Torres Strait Islander, Culturally and Linguistically Diverse (CALD), and low socio-economic status (low SES) people.	Online cross-sectional survey about consumer use, understanding and trust in the HSR and evaluate performance of HSR education campaign.	HSR only (on product)	Low Rated moderately on research design. Rated poorly across almost all areas of research procedures. Concerns with the overall execution.
Pollinate Research (2016)	2016	Quantitative	N = 1,007 Nationally representative sample of	Online cross-sectional survey about consumer use, understanding and trust in the HSR and	HSR only (on product)	Low Rated moderately on research design.

Study (authors/ date)	Date data collected	Study type	Population	Design	HSR stimuli shown	Quality Assessment
			main/joint grocery buyers aged 18 years and older, with a population representative of Aboriginal and Torres Strait Islander, Culturally and Linguistically Diverse (CALD), and low socio-economic status (low SES) people.	evaluate performance of HSR education campaign.		Rated poorly across almost all areas of research procedures. Concerns with the overall execution.
Pollinate Research (2017)	2017	Quantitative	N = 1,052 Nationally representative sample of main/joint grocery buyers aged 18 years and older, with a population representative of Aboriginal and Torres Strait Islander, Culturally and Linguistically Diverse (CALD), and low socio-economic status (low SES) people.	Online cross-sectional survey about consumer use, understanding and trust in the HSR and evaluate performance of HSR education campaign.	HSR only (on product)	Low Rated moderately on research design. Rated poorly across almost all areas of research procedures. Concerns with the overall execution.
Pulker et al. (2019)	15mth after HSR implemen tation	Qualitative	N = 37 Four fathers and 33 mothers of children aged 2–8 years, main household food shopper.	Five 90-min focus groups; semi-structured guide directed focus group discussions.	HSR only (on product) and HSR with tail (on product)	High Rated moderately on sample diversity, recruitment reporting, procedural clarity, and tool development. Rated poorly on sample representativeness and analytical depth. No major concerns with the overall execution of the study.

Study (authors/ date)	Date data collected	Study type	Population	Design	HSR stimuli shown	Quality Assessment
Reilly et al. (2018)	2016	Qualitative/ Quantitative	N = 91 Primary school canteen managers from the Hunter New England (HNE) region of NSW.	RCT Telephone interview which of 12 common food products sold in school canteens they would sell. Both groups received product name and brand information. The intervention group also received information regarding the nutritional rating of products.	No stimuli given	Medium Rated moderately on recruitment reporting, procedural clarity, and statistical alignment with study design. Rated poorly on sample representativeness and measurement tool validation. Some concerns with overall execution.
Riley et al. (2016)	Sep to Dec 2015	Quantitative	N = 3,005 1816 women, and 1189 men; with 626 aged 15-34 years; 850 aged 35-54 yrs; and 1529 aged above 55 years. Respondents living in the metropolitan area numbered 2120, with 885 living in a rural area.	Face to face cross-sectional survey about how various elements of food labels are used to inform their purchase decisions.	No stimuli given	High Rated high on sample design and representativeness. Rated moderately on recruitment reporting, procedural clarity, and data analysis. Rated poorly on measurement tool validation. Some concerns with overall execution due to the very limited assessment of the HSR.
Russell et al. (2017)	Not stated	Quantitative	N = 520 Parents (53% male) with a child aged between five and eleven years. Median age of 37 years. Sole or joint decision maker for purchasing breakfast cereal for their child.	Discrete Choice Experiment Participants asked to select the product they most preferred and the product that they least preferred from the four cereal options available (choosing for their children).	HSR only (on product)	High Rated high on analytical approach. Rated moderately on procedural clarity and partial use of validated tools. Rated poorly on sample representativeness and recruitment reporting.

Study (authors/ date)	Date data collected	Study type	Population	Design	HSR stimuli shown	Quality Assessment
						No major concerns with the conduct of the study.
Talati et al. (2019a)	2018	Quantitative	N = 12,015 from 12 countries including 987 Australians. Quotas were applied so that the sample was evenly split according to gender, age (within the following brackets: 18–30 years, 31–50 years, >50 years) and income level (low, medium and high).	RCT repeated measures design Respondents were shown foods of varying nutritional quality (with no label on package) and selected which they would be most likely to purchase. The same choice sets were then shown again with 1 of 5 randomly allocated labels on package (HSR, Multiple Traffic Lights, Nutri-Score, Reference Intakes, or Warning Label).	HSR with tail (on product)	Medium Rated high on large international sample size, standardised stimuli, and appropriate statistical approach. Rated poorly on recruitment detail, tool validation, reporting of subgroup effects, and result presentation clarity. No major concerns with overall execution.
Talati et al. (2016a)	Not stated	Qualitative	N = 85 including 50 adults (27 males and 23 females) and 35 children (18 males and 17 females). Groups ranged in size from seven to 10 participants and were segmented according to gender (male, female) and age (10–13, 14–17, 18–25, 26–45, 46+ years).	Ten focus explored adults' and children's reactions when presented with both a FOPL (the Daily Intake Guide, Multiple Traffic Lights, or the Health Star Rating) and a health claim (nutrient content, general-level-, or high-level). Conducted in Perth, Western Australia.	HSR with tail (on product and not on product)	Medium Rated moderately on procedural reporting. Rated poorly on recruitment transparency, tool development, and analytical rigour. Some concerns with overall execution.
Talati et al. (2016b) – Same as	Not stated	Qualitative	N = 85 including 50 adults (27 males and 23 females) and 35 children	Ten focus explored adults' and children's reactions when presented with both a FOPL (the Daily Intake	HSR with tail (not on product	Medium Rated high on statistical analysis methods.

Study (authors/ date)	Date data collected	Study type	Population	Design	HSR stimuli shown	Quality Assessment
above study			(18 males and 17 females). Groups ranged in size from seven to 10 participants and were segmented according to gender (male, female) and age (10–13, 14–17, 18–25, 26–45, 46+ years).	Guide, Multiple Traffic Lights, or the Health Star Rating) and a health claim (nutrient content, general-level-, or high-level). Conducted in Perth, Western Australia.	and on product)	Rated moderately on sample variation and procedural reporting. Rated poorly on recruitment and measurement tool development. Some concerns with overall execution.
Thomas, Seenivasan & Wang (2021a)	Not stated	Quantitative	N = 171 (42% male) students in a marketing course	2x2 between-subjects experimental design (high salt/high fat (HSHF), 2 stars; low salt/low fat (LSLF), 3.5 stars); (low salt/high fat (LSHF), 3 stars; and high salt/low fat (HSLF), 2.5 stars.	HSR only (on product, next to NIP)	Low Rated moderately on procedural clarity and partial use of validated subscales. Rated poorly on sample representativeness, recruitment transparency, and appropriateness of stimuli and measurement tools. Major concerns with overall execution.
Thomas, Seenivasan & Wang (2021b)	Not stated	Quantitative	N = 820 respondents (52% female) from an online panel	Experimental design with two between-subjects conditions (HSR: present and absent).	HSR only (on product)	Low Rated moderately on research procedure description and use of validated subscales. Rated poorly on sample representativeness, recruitment transparency, measurement alignment with policy context, and outcome definition. Major concerns with overall execution.
Thomas, Seenivasan	Not stated	Quantitative	N = 103 (28.2% male) students in a marketing course	Eye tracking study	HSR only (on product)	Low

Study (authors/ date)	Date data collected	Study type	Population	Design	HSR stimuli shown	Quality Assessment
& Wang (2021c)				Eye-fixation data were used to estimate the formation of heuristic rules for choice sets of products with and without HSR and different nutrient content.		Rated moderately on procedural description and use of appropriate statistical methods. Rated poorly on sample representativeness, recruitment transparency, tool validation, and measurement alignment with intended policy context. Major concerns with overall execution.
Thomas, Seenivasan & Wang (2021d)	2014 to 2016	Quantitative	Australian Nielsen Home scan panel data set. Data set covers 6 months of data before the HSR introduction and 30 months of data after its Introduction.	Neilson Home scan data matched to food switch data to identify whether the healthiness of foods purchased by consumers improved after the implementation of the HSR system in Australia.	HSR only (not on product)	Low Rated moderately on analysis methods. Rated poorly on sampling rationale, recruitment procedures, representativeness, procedural transparency, and tool validation. Major concerns with overall execution.

New Zealand Studies

Study (authors/date)	Date data collected	Study type	Population	Design	HSR stimuli shown	Quality Assessment
Bablani et al. (2022)	2013 to 2019	Quantitative	N = 2,500 households from the Nielsen HomeScan continuous open cohort panel, 2013–2019. Major urban and secondary urban sites in NZ (sampled from 92% of NZ population).	Secondary data analysis Fixed effects analyses were used to estimate the association of HSR with product and nutrient purchasing	All forms (on product)	High Rated high on large sample size and use of appropriate, robust analysis methods. Rated moderately on recruitment details, representativeness, and data collection tools. No major concerns with overall execution.
Colmar Brunton (2016a)	Oct and Nov 2015	Quantitative	N = 1,678 shoppers including: Priority populations with children <14 years Low income (n = 324) Māori (n = 300) Pacific (n = 311). General population (n = 743)	Online cross-sectional survey	HSR only (on product) and HSR with tail (on product)	Medium Rated high on sample description and data collection procedures. Rated moderately on recruitment reporting, tools, and analysis. Rated poorly on ethics reporting. No major concerns with overall execution.
Colmar Brunton (2016b)	Sep and Oct 2016	Quantitative	N = 1,658 shoppers including: Priority populations with children <14 years Low income (n = 309) Māori (n = 310) Pacific (n = 303) General population (n = 736)	Online cross-sectional survey	HSR only (on product) and HSR with tail (on product)	Medium Rated high on data collection transparency and reporting clarity. Rated moderately on analysis approach, recruitment detail, and validity of measures. Rated poorly on ethics reporting. No major concerns with overall execution.

Study (authors/date)	Date data collected	Study type	Population	Design	HSR stimuli shown	Quality Assessment
Colmar Brunton (2018)	Feb and Mar 2018	Quantitative	N = 1,645 shoppers including: Priority populations with children <14 years Low income (n = 316) Māori (n = 301) Pacific (n = 307) General population (n = 721)	Online cross-sectional survey	HSR only (on product) and HSR with tail (on product)	Medium Rated high on data collection transparency and reporting clarity. Rated moderately on analysis approach, recruitment detail, and validity of measures. Rated poorly on ethics reporting. No major concerns with overall execution.
Hamlin and McNeill (2018)	2016	Quantitative	N = 2,600 Dunedin n =1.000 Christchurch n= 1.600 Recruited by qualified intercept by observing purchase of a breakfast cereal.	2 x2 factorial design with 2 product treatment levels: one product with a high and one product with a low nutritional status, and two HSR FoP treatment levels: FoP reflecting nutritional status present, and FoP absent	HSR with tail (on product)	Medium Rated moderately on recruitment reporting. Rated poorly on sample representativeness, measurement tools, procedural clarity, and analytic depth. Some concerns with overall execution.
Hamlin and McNeill (2016)	2014	Quantitative	N = 1,200 consumers in Dunedin Recruited by qualified intercept. Limited data collected/reported on consumers.	2x2 factorial design was used with two levels of cold cereal product nutritional status (high, five star/low, two star) and two levels of the Health Star Rating label (present/absent).	HSR with tail (on product)	Medium Rated moderately on sample representativeness and recruitment reporting. Rated poorly on measurement tool validation, procedural clarity, and analytic depth. Some concerns with overall execution.
The Navigators (2024)	2023	Quantitative	N = 1,602 aged 15 years and older including:	Online cross-sectional survey exploring consumers' attitudes, knowledge and	No stimuli given	Medium Rated moderately on representativeness, data collection clarity, and

Study (authors/date)	Date data collected	Study type	Population	Design	HSR stimuli shown	Quality Assessment
			General public (n = 1,040) Māori (n = 414) Pasifika (n = 403) Asian (n = 172)[7]	behaviours around food safety and suitability.		consideration of validity through cognitive testing and piloting. Rated poorly on sample size justification, recruitment detail, and ethics reporting. No major concerns with overall execution.
Ni Mhurchu et al. (2017)	2014 to 2015	Quantitative	n = 1,357 participants were 18 years or older, owned a smartphone, were the main household shopper and shopped at a supermarket 1 or more times per week.	5-wk, 3-arm, parallel-group RCT; Eligible participants were randomly assigned (1:1:1) to receive either TLL or HSR or control (NIP). Smartphone technology allowed participants to scan barcodes of packaged foods and to receive allocated labels on their smartphone screens.	All forms (on product)	High Rated high on statistical analysis aligned and sample size planning. Rated moderately on sample representativeness, recruitment reporting, procedural clarity, and measurement tools. No major concerns with overall execution.
Stuthridge et al. (2022)	2020	Qualitative	N = 49 over the age of 25 years; 10 recruited for interviews and 39 for focus groups. Requirement to be the primary food shopper in the household and live in Dunedin, Wellington or Nelson. Quotas for holding a Community Services card, having children, and from Māori or Pasifika households.	Semi-structured in-depth online interviews and in-person focus groups.	HSR only (on product) and HSR with tail (on product)	High Rated high on efforts to capture demographic and ethnic diversity relevant to the NZ context. Rated moderately on recruitment reporting, procedural clarity, tool development, and analysis depth. No major concerns with overall execution.

[7] Respondents could identify with more than one ethnicity.

Study (authors/date)	Date data collected	Study type	Population	Design	HSR stimuli shown	Quality Assessment
TNS New Zealand (2015)	Sep 2015	Qualitative	N = 48 7 focus groups: 2x Māori low income, 1x Māori medium income, 1x Pacific low income, 1x Pacific, medium income, 1x other ethnicities low income, 1x other ethnicities medium income Were the main or joint shopper for their household with children under 14 years in the household.	Focus groups Face to face	HSR only (on product) and HSR with tail (on product)	High Rated high on procedure description and research setting. Rated moderately on some aspects of recruitment data and analysis approach. No major concerns with overall execution.

Australian and New Zealand Studies

Study (authors/ date)	Date data collected	Study type	Population	Design	HSR stimuli shown	Quality
FSANZ (2025a)	2024	Quantitative	AU N = 1,231 NZ N = 884 representative by age, gender, and location.	Online cross-sectional survey – CIT 2024	HSR only	High Rated high on large sample size, use of appropriate analysis methods, representativeness, and data collection tools. Rated poorly on recruitment details, use of a theoretical framework, and limitations. No major concerns with overall execution.
FSANZ (2025b)	2025	Qualitative/ Quantitative	AU N = 1,554 NZ N = 696 representative by age, gender, and location. Also representative of Māori and Pacific Peoples in NZ, and Aboriginal and Torres Strait Islanders in Australia.	Online cross-sectional survey – HSR Survey 2024	HSR only HSR + energy HSR + tail (on and not on products)	High Rated high across most areas. No major concerns with overall execution.
FSANZ (2024)	2023	Quantitative	AU N = 1,237 NZ N = 810 representative by age, gender, and location.	Online cross-sectional survey – CIT 2023	HSR only	High Rated high on large sample size, use of appropriate analysis methods, representativeness, and data collection tools. Rated poorly on recruitment details, and use of a theoretical framework. No major concerns with overall execution.
Hallak et al. (2021)	2019	Quantitative	AU N = 808 NZ N = 213 representative by age, gender, and region. Participants who purchase/ eat out at cafes, restaurants, pubs, and takeaways were	Online cross-sectional survey	Not stated if stimuli was used or not	Low Rated poorly across most areas. Concerns with the overall execution and strength of the conclusions.

Study (authors/ date)	Date data collected	Study type	Population	Design	HSR stimuli shown	Quality
			included in the sample frame.			
Pettigrew et al. (2020b)	Not stated	Quantitative	N = 7,545 from seven countries including 1,033 Australians and 1,090 New Zealanders Quotas applied for equal distribution by gender and three age categories (18–34, 35–54, 55+ years), and with at least two-thirds of the sample in the low- and middle-income categories	RCT Online choice tasks requiring them to select a preferred breakfast cereal and then nominate the healthiest cereal	HSR only (on product) and HSR with tail (on product)	Medium Rated high on critical discussion of design limitations, label stimulus clarity, and large international sample.