**Supplemental Table 1** Quality Assessment Scheme

|  |  |
| --- | --- |
| Design & methods | Points |
| Study setting & Outcome | Real-world measuring energy consumption. | 3 |
| Experimental measuring energy consumption. | 2 |
| Real-world measuring energy ordered (verified by purchase receipt). | 2 |
| Experimental measuring energy ordered (verified by purchase record). | 1 |
| Real-world measuring energy ordered or consumed (based on self-reported recall). | 1 |
| Experimental measuring energy selected. | 0 |
| Extent of ML | Complete, all items on the menu were labelled and reference values are specifically included. | 2 |
| Complete, all items on the menu were labelled but the presence of reference values is unknown or absent. | 1 |
| Partial – selected items on the menu were labelled or not reported. | 0 |
| ML “noticing” rate  | High (≥ 70%) | 1 |
| Normal (50-70%) | 1 |
| Not reported / Low (≤ 50%) | 0 |
| Sample size | Large sample (> 150 per group)  | 2 |
| Medium (51-150 per group) | 1 |
| Small sample (≤ 50 per group) | 0 |
| Randomization (for experimental settings)OR Suitable case-control match (for real-world settings) | Participants were randomly allocated to intervention or control groups. | 1 |
| The case-control site/outlet was a suitable match for the real-world labelled site/outlet. | 1 |
| Non-randomization, or non-suitable case-control match, or no description. | 0 |
| Study population  | Analysis shows the study populations are similar according to socio-demographic characteristics or differences were adjusted for in further analyses. | 1 |
| Statistically significant (P<0.01) socio-demographic differences between study populations, which were not adjusted for in further analyses, or lack of study populations’ comparison. | 0 |
| Blinding | Blinded – blinding of investigators, analysts, or participants to the intervention allocation to the extent possible. | 1 |
| No blinding or process not described. | 0 |
| 0-3 points – Weak | 4-7 points – Fair | 8-11 points – Good  |

Model developed with inspiration from Krieger & Saelens (2013); Sinclair, Cooper & Mansfield (2014), Swartz, Braxton & Viera (2011).

Supplemental Table 2 Quality assessment scheme, results. Studies are grouped according to the nature of the outcome.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Reference | Study setting | Outcome type(energy) | Extent of ML& Reference Values (RV) | ML noticing rate\* | Sample size† | Randomisation or case-control match | Population‡ | Blinding | Rating |
| Hammond *et al.* 2013 (44) | Experimental  | Ordered & consumed | Complete | High | Large | Randomised | Similar  | Blinded | Good |
| James *et al.* 2014 (43) | Experimental  | Ordered & consumed | Complete + RV | High | Medium | Randomised | Adjusted | Blinded | Good |
| Vanderlee & Hammond 2014 (21) | Real-world | Ordered & consumed | Partial | High | Large  | No case-control match | Adjusted | Not described | Fair |
| Auchincloss *et al.* 2013 (22) | Real-world | Ordered | Complete§ | High | Large | Case control, not matched | Adjusted | Not described | Fair |
| Elbel *et al.* 2013 (46) | Real-world | Ordered | Complete | Low | Large | Case-control match | Adjusted | Not described | Fair |
| Krieger *et al.* 2013 (45) | Real-world  | Ordered | Complete(90%) | Food - Normal Coffee - Low  | Large | Case-control match | Similar (except for age) | Not described | Good |
| Brissette *et al.* 2013 (47) | Real-world | Ordered | Complete | Not reported | Large | Case-control match | Different | Blinded | Fair |
| Ellison *et al.* 2013 (39) | Real-world | Ordered | Partial | Not reported | Small | Randomised | Not reported | Not described | Weak |
| Holmes *et al.* 2013 (38) | Real world | Ordered | Partial | Not reported | Large | Within group  | Similar  | Not described | Fair |
| Dodds *et al.* 2013 (48) | Telephone survey  | Selected | Complete + RV | High | Medium | Randomised | Similar | Blinded | Fair |
| Liu *et al.* 2012 (16) | Online survey | Selected | Complete + RV | High | Medium | Randomised | Adjusted | Blinded | Fair |
| Dowray *et al.* 2013 (50) | Online survey  | Selected | Complete | Not reported | Large | Randomised | Similar  | Not described | Fair |
| Morley *et al.* 2013 (49) | Online survey | Selected | Complete + RV | Not reported | Large | Randomised | Similar | Not described | Fair |
| Pang & Hammond 2013 (4) | Experimental | Selected | Complete + RV | Not reported | Small | Randomised | Different | Not described | Weak |
| Roseman *et al.* 2013 (40) | Experimental | Selected | Complete | Not reported | Medium | Randomised | Not reported | Not described | Weak |

\* Low ≤ 50%, Normal 50-70%, High > 70%

† Small ≤ 50 per group, Medium = 51-150 per group, Large > 150 per group.

‡ Different = significant socio-demographic differences between groups, not adjusted for during data analysis; Similar = no significant socio-demographic differences between groups or across waves; Adjusted = significant socio-demographic differences between groups were adjusted for during data analysis.

§ 60% of soft drinks were not labelled but they were analysed separately.