SUPERLIST

Research Framework

4

June 2023, Version 1.4, Copyright Questionmark
Authors: Gustaaf Haan, Willem van Engen, Deborah Winkel, Charlotte Linnebank

Content

Background

The Superlist research programme analyses supermarkets' role in the food system. The programme alternatingly looks into one of four themes: health, environment, human rights and animal welfare. Every study results in a comparative ranking of supermarkets, to show the extent to which they contribute to a responsible food environment.
Prior to each thematic study, Questionmark devises research criteria according to which supermarkets will be compared. These criteria are based on the latest scientific findings in areas such as sustainability, health and consumer behaviour. In consultation with scientists and experts, these findings are translated into unambiguous, measurable criteria.
In order to give all supermarkets equal opportunity to optimise their performance, we discuss our criteria with them well in advance. The criteria are also made available to the public on Questionmark's website.

Our research approach and underlying principles are laid down in this document. It provides the framework for the development of research criteria for individual thematic studies, ensuring that publications in the Superlist research programme always provides a fair, independent and useful picture of the differences between supermarkets.

Questionmark

Questionmark strives for a world in which people live healthy lives, on a healthy planet, with respect for all people and animals. For detailed information about Questionmark's mission, please see the website.

Financial support

The development of the Superlist approach was co-financed by the DOEN Foundation and the Questionmark Foundation. The translation of the research framework from Dutch to English was financed by EU LIFE. Individual thematic study may be supported financially by other funds or by civil society organisations. All financial contributions are mentioned in the publications related to the study.

Objective

Objective and central criterion

Supermarkets[1] have an influence on what people in the Netherlands eat and drink: in Northwest Europe we purchase roughly three quarters of our food in supermarkets, which makes supermarkets key players in many supply chains. Superlist helps and encourages supermarkets to leverage their influence over the food system to make diets healthier and more sustainable.

The central criterion we use to compare supermarkets is the following question:

To what extent does the supermarket as an environment in which people make their daily food choices, promote a sustainable and healthy diet?

We exclude other ways in which supermarkets can take social responsibility, such as sustainable business operations or partnerships with charity organisations.

Themes

Supermarkets' efforts are thus analysed through four thematic lenses:

- Health
- Environment
- Human rights
- Animal welfare

We sometimes use 'responsible food' and ‘sustainable food’ as an umbrella term for healthy and sustainable food, and interpret 'sustainability' as a combination of three themes: environment, human rights and animal welfare.

[1] We use the term 'supermarket' to refer to a chain of supermarkets known to consumers under a single brand name. When this risks causing confusion, we use the word 'branch' to refer to individual stores.

Theory of change

Superlist contributes to a more responsible food system in two different ways.

1. Superlist's primary aim is to bring about change at supermarkets. Supermarkets are given suggestions of interventions they could implement to promote sustainability and health, illustrated by best practices. This offers them the opportunity to stand out as a frontrunner or risk being seen as a laggard (which can be just as much of an incentive!).
2. Superlist's secondary aim is to contribute to the public debate about healthy and sustainable food by, for example, raising awareness of certain problems and solutions in the food system and underpinning arguments with facts and figures.

Research principles

Our research approach is based on five main principles.

Relevance
Our research is designed to help supermarkets take actions that will substantially contribute to making our food system healthier and/or more sustainable.

Independence
Our research method is free from bias against or in favour of individual supermarkets. We avoid all influence of commercial interest or semblance thereof.

Level playing field
We assess each supermarket in scope of the study without prejudice and we make no distinction between supermarkets other than in line with the purpose of this project.

Transparency
The research methods we employ to compare supermarkets are available for the public to read. Given the raw data that Questionmark has gathered, anyone should be able to go through all calculations and verify our results.

Fair hearing
We inform the supermarkets of our comparative criteria well in advance. We give them the opportunity to make suggestions to improve the research approach. Supermarkets are also given the opportunity to review and correct the data we collect and to comment on our eventual findings.

Formulating comparative criteria

Our comparison is based on specific criteria for each theme, which are laid down in separate documents. All criteria are described at three different levels:

- The **issues**, that is: problems in the food system that require solutions to which supermarkets could make a substantial contribution.
- The possible **actions / interventions** by the supermarket to address the issue. Suggested interventions can be formulated in terms of change in assortment, promotions, shop layout or policy
- The **indicators** that make the interventions measurable. Each indicator has a weighting that expresses the relative importance of the indicator for the theme as a whole.

All comparative criteria are formulated according to the following conditions, which are based on the above mentioned research principles.

Conditions

1) The comparative criteria address the most crucial problems in the food system, which require solutions that supermarkets could make a significant contribution to.

2) The desired interventions are relevant to help solve the problems, at least during the next five years.

3) The indicators measure how supermarkets perform in terms of taking the suggested actions; a supermarket that scores well on all indicators for a theme is contributing relatively well to making our food system more responsible in relation to that theme.

4) We expect that there will be a marked difference between supermarkets in the Dutch market in terms of aggregated scores per theme, or that such a difference will arise soon after indicators are introduced.

5) The selected issues appeal to a broad audience.

Formulating indicators

The research principles lead to the following conditions for indicators.

Conditions

1) Supermarkets can only improve their score on a certain indicator by taking actions relevant to the theme that the indicator relates to.

2) The link between the indicator and the theme it relates to is preferably scientifically substantiated, or at least highly plausible.

3) All supermarkets have ample opportunity to improve their score on an indicator. While some supermarkets may lag further behind due to choices they made in the past, there are no external factors that make progress more difficult for one supermarket than another.

4) The indicator is formulated in such a way that there is no upper limit to the score supermarkets can achieve (at least during the coming five years); there is always room to improve one's score.

5) The indicator is measurable regardless of the supermarket's cooperation, and measuring it does not take an excessive amount of time or money.

6) The indicator is unambiguous. If necessary, the terms it uses are supplied with a clear definition.

7) Where possible, the indicator contributes to the (international) harmonisation of social requirements and standards for health and sustainability.

Determining relative weights

Every indicator is assigned a weight of 0.5, 1 or 1.5 times the score. An indicator's weight is determined by the urgency of the suggested intervention and the issue this intervention seeks to address. Weights are discussed with the Scientific Board and relevant civil society organisations, based on the following condition:

1) When relative weights are assigned to a theme's indicators, only the importance of the issues and interventions for society may factor into the decision.

Changing indicators over time

For each replication of a study, the research methodology may be revised. Criteria may be adjusted to developments in the market or in society, to new scientific insights or to local circumstances. In adjusting the criteria, the

conditions as formulated above are used as guiding principles. On top of that, the following condition holds for each change in indicators:

1) Whenever possible, the adjusted indicator is backwards compatible with the original indicator in the sense that its results allow for comparison with results of earlier versions.

Input from civil society

For an optimal impact of Superlist, garnering broad support for the research method is a must. That is why we gather input from various stakeholders when developing our research method.

Scientific Advisory Board

A Scientific Board (or the Board, for short) was set up to advise Questionmark on the development of its research method, which includes both this Research Framework and the Comparative Criteria. Questionmark consults the Board to ensure its Superlist research is in line with the latest scientific findings regarding issues such as:

- theory of change,
- relevance of the proposed comparative criteria,
- robustness of the proposed indicators and data collection methods.

If the data can be interpreted in different ways, or if circumstances force us to deviate from our research method, our research team will always discuss the matter with (members of) the Board.
Board members provide advice as individuals, which means the Board does not need to be unanimous and Questionmark may decide not to follow its advice. Should Questionmark choose to ignore a suggestion that the majority of the Board agreed upon, it will clarify its decision in the final report published for the theme in question.
The Board's way of working is described in a separate document, which is available on www.superlijst.org, which also lists the current members of the Scientific Board.

Cooperation with civil society organisations

Questionmark invites civil society organisations to help assign priorities to different food system problems that the Superlist project will look into.
Civil society organisations can propose issues to be addressed, corresponding interventions and indicators and suggest a weighting. As experts on a certain theme, organisations each have a voice as individual advisors.

The names of these organisations will be announced in the comparative criteria document of each theme. Organisations that opt to collaborate more

closely with Superlist (hereafter sometimes referred to as 'partner NGOs') will also contribute financially to the research.

Consulting supermarkets

We consult the supermarkets at four different moments during the research.

1) When we develop comparative criteria, we will ask supermarkets to suggest improvements for the proposed indicators. Questionmark may decide to adopt supermarkets' suggestions, as long as they are in line with this research framework and as long as they contribute to the objective of Superlist. Questionmark will always explicitly discuss any such edits with the Board.
2) At the start of the research, or no less than 8 weeks before its conclusion, we will inform supermarkets of the definitive comparative criteria. At that time, we will also announce the cut-off date, which marks the end of the period during which data is collected. Up until the cut-off date, supermarkets can inform the research team of relevant policy changes or changes to their range.
3) After the cut-off date, supermarkets will be given an overview of the main data the researchers will use, with the request to correct any possible errors in the data. Submitted corrections will first be reviewed by Questionmark.
4) Finally, supermarkets will also be given the chance to react to the ranking. Substantive reactions may be included in the final report or in the press release.

Data collection & scoring

Scope

Superlist focuses on the largest supermarket chains (in terms of turnover or the number of branches) that together hold a market share of at least 80%. The sample may include supermarkets that specifically claim to promote health or sustainability, provided they have a nationwide network of branches. Which supermarkets are or aren't included in scope is explicitly stated in the comparative criteria for each theme.

Product data

Comparing the ranges of different supermarkets requires up-to-date information about things such as the composition and origin of the products the supermarket sells. Whenever possible, we obtain such data from digital sources. If a supermarket has a webshop, product data is collected via its webshop. In the absence of a webshop, we visit a supermarket's (major) branches to survey its range. We photograph products that are relevant to our research and digitise their product data, to ensure all product data we collect is eventually stored in one and the same database. This data is then combined with data from other sources, such as data obtained from the industry at large via the GS1 data exchange platform, and any available additional data. Supermarkets and brand owners can also provide us with product data themselves to make sure Questionmark has the most up-to-date data. During every analysis, we perform both automatic and manual checks to make sure the quality of our data is sufficient for the purposes of this project.

Promotions

To get a picture of supermarkets' promotion policy, we analyse their promotions run during the research period. Depending on local circumstances we may decide to use as a source:

- The promotions leaflets that supermarkets publish, either physically or on their website.
- All promotions run in the online store.

Policy and objectives

Supermarkets' policies are analysed by reviewing supermarkets' websites and, where applicable, the website of their holding company or purchasing organisation. Questionmark endeavours to review all parts of a supermarket's

website, but also asks supermarkets to point out web pages they themselves feel are most relevant. Online annual (CSR) reports will always be taken into account.
Policy changes will be taken into account up until the cut-off date, which is communicated to all supermarkets in advance. On the cut-off date, we will make a local copy of all web pages we feel contain relevant information.

Spot checks

Suggested interventions may not only concern products or promotions, but also a supermarket's layout or interior design. Visiting every single branch of a supermarket chain is beyond the scope of this research. To still get an idea of the extent to which a supermarket is taking this kind of action, we review policy documents and conduct spot checks at some of the supermarket's branches. Indicators for which we do so are marked with 'SPOT CHECK', with research then conducted as follows.

1) We investigate whether or not a supermarket has formulated a policy that all branches should adhere to.
2) We conduct spot checks at branches across the entire region in scope. Spot checks are also conducted at supermarkets that do not have an official policy, as they could still have implemented the suggested intervention. Spot checks are conducted according to the following rules of thumb:

    - No more than one branch of the same chain in one and the same location.
    - For each chain, we visit at least two branches in a village or small town and two in a city or urban area.
    - For each chain, we visit branches in at least three different provinces or districts.
    - We visit regular branches, not smaller branches (at for example railway or petrol stations) or branches with an experimental or unusual formula.

Analysis and scoring

Policy goals
For some themes, the first step a supermarket could take to improve its ranking is to formulate short-term goals. 'Short term' does not necessarily mean the same thing to all supermarkets. We will proceed as follows to be able to compare goals with differing deadlines. We will consider all goals that supermarkets are seeking to achieve within the next five years. Goals predicted to take more than 1 year to achieve will be linearly interpolated for the years in between. If, for example, a supermarket aims to achieve a 100%

improvement over the coming two years, we assume the target for the first of those two years will be a 50% improvement.

Reliability factor

Any score on a policy-goal indicator is always multiplied by a reliability factor:

- If there was no previous goal — factor 0.5
- If the previous goal was fully achieved — factor 1
- If 50% or more of the previous goal was achieved — factor = %
- If less than 50% of the previous goal was achieved — factor 0.1

If a supermarket has not reported on the extent to which it achieved a previous goal, we will assume less than 50% of that goal was obtained.

Interim actions

As comparative criteria are developed, stakeholders can suggest 'interim actions' along the way to implementing the desired interventions. They may do so because, for example, a certain intervention is important but hard for supermarkets to implement. These interim actions will be easier to achieve but may be less effective or ambitious.
The score for an interim action is always limited or maximised relative to the possible score for the desired intervention. The comparative criteria explicitly
indicate when interim actions are taken into account, what they may be and how they are scored.

Reliability factor via spot checks

In the ideal situation, policy and practice are in agreement. The results of a spot check test can be used to estimate the value of a supermarket's policy. The table below shows how the 'practice factor' can be included in the valuation of policy.

Table 1. Factor by which a supermarket's score is multiplied		Number of branches where action is taken from a spot check of 5 branches		
		0 branches	**1-4 branches**	**5 branches**
Action part of published policy?	**Yes**	Factor: 0	Factor: 0	Factor: 1
	No	Factor: 0	Factor: 0,25	Factor: 0,5

Table 1. Determination of the 'policy practice factor' for policy indicators

Examples
The official policy of supermarket chain X is that all branches should implement intervention z. A random spot check of 5 branches reveals that intervention z has not been implemented in 2 of these branches, which thus do not pass the test. Supermarket chain X' score on this indicator is therefore multiplied by factor 0, which means X is awarded a score of 0.

Supermarket chain Y does not mention intervention z in any official policy documents. However, a spot check of 5 branches of Y reveals that the intervention has been implemented at all branches that were visited. The score of a supermarket chain Y is multiplied by factor 0.5, which halves its score.

Private label factor
A supermarket is responsible for its entire assortment. In practice, policy or reporting is sometimes limited to private label products. Some supermarkets sell almost exclusively private labels, for other supermarkets it is only a small part of the assortment.

A private label factor is used to make a statement about private labels comparable. For policy and reporting indicators, the private label factor used is equal to the share of private label products in total sales, as reported by the supermarket.

If the supermarkets do not report on the share of private label products in sales figures or assortment, Questionmark determines the private label factor based on the share of private label products for the supermarket in question in Questionmark's database, according to the table below.

Share of private label	Private label factor
0% - <30%	0.2
≥30% - <60%	0.4
≥60% - 100%	0.6

The private label factors that are used are published in the comparison criteria for each Superlist.

Scaling

An indicator measures each supermarket's performance and expresses it with a single metric. Sometimes, this will be a number between 0 and 100. At other times, it will be a number that, in theory, is infinitely large or small. Either way, the metric needs to be translated into a score on a scale of 0 to 100, with 0 signaling a very poor performance and 100 a stellar performance.
There are different ways to 'scale' metrics, with some approaches being better suited to certain indicators than others. In each theme's comparative criteria, we mention the scaling approach we selected for each indicator. Below, we briefly describe the different approaches.

No scaling

In this approach, the maximum score (100) is equated with the maximum that supermarkets could in theory achieve. For example, if intervention X is desirable for as many products as possible, the intervention applying to a supermarket's entire range would be the maximum result. If it applies to only 80% of the range, the supermarket will score 80 points on that indicator. It could well be the case that no supermarket scores 100 and no supermarket scores 0. Not scaling a metric is a good approach for indicators on which supermarkets score wildly different scores, because the scale will automatically illustrate these vast differences.

Scaling fully

In this approach, the actual results determine how scores are scaled. Whichever supermarket turns out to be the poorest performer will get a score of 0. The best-performing supermarket gets a score of 100. Other supermarkets receive scores that reflect their performance in relation to the best- and worst-performing supermarkets. If another supermarket scores even better in the next study (a few years later), the scale is automatically adjusted: the new frontrunner gets a score of 100 while the other supermarkets' scores drop (unless their policies have equally improved). This

approach to scaling is a good fit for indicators that result in more subtle differences.

Suppose all supermarkets have implemented intervention Y to 80% of their range – but 100% needs to be achieved to create a responsible food environment. And suppose that one supermarket managed to achieve 87%. Without scaling, the 7 percentage-points difference between this frontrunner and the other supermarkets would be barely visible. By fully scaling results, the supermarket with a coverage of 87% scores 100, while a supermarket that scores 80% (just like most of its peers) gets a score of 0. This approach zooms into the difference between supermarkets and renders those differences more visible.

Fixed limits

In this approach, we predetermine an upper and lower limit for supermarkets' expected performances, even though results exceeding those limits are theoretically possible. Predetermining limits may be necessary when, for example, a metric does not fall between 0 and 100 but is instead a number between 0 and infinite. In this case, the predetermined upper and lower limits determine our scale.

A supermarket that, for example, scores a number that lies right in the middle between those two limits will get a score of 50. Supermarkets that achieve the upper limit get a score of 100, but so do all supermarkets that perform even better. The advantage of this way of scaling is that it allows the upper limit to serve as a target to aim for – whereas with the 'scaling fully' approach, any supermarket that outperforms its peers even slightly immediately gets a score of 100.

Flexible limits

This approach resembles the 'fixed limits' approach, with the difference being that if one supermarket exceeds the predetermined upper limit, the scale is adjusted accordingly. Other supermarkets whose performance is right on the upper limit then automatically receive a score lower than 100. The same applies to the lower limit: the scale is adjusted downwards if a supermarket's performance is worse than the lower limit. This approach has the same advantages as the 'fixed limits' approach, but offers greater flexibility when there is no advance knowledge of how well supermarkets will perform on an indicator.

Minor differences

If there are only minor differences in two or more supermarkets' performances on an indicator, the quality of the data for that indicator may

need to be reviewed. If the error margin for that specific indicator is of the same order as the differences found between supermarkets, Questionmark may decide to equate supermarkets' scores or disregard the indicator in question.

Score visualisation

Visualising the results serves to highlight the differences in performance between supermarkets, which both rewards frontrunners and serves as an incitement for laggards. Here is how we visualise results:

Per supermarket, we calculate a final score, which consists of the weighted average of the supermarket's scores on all indicators. The order of supermarkets in the ranking is determined by these final scores, with the supermarket with the highest final score topping the list.

Each supermarket's final score is reflected in the length of a horizontal bar, with the bar's different colours representing the supermarket's partial scores for each issue. Final scores are not quoted as numbers but are mentioned instead in the accompanying background report.

In the final-score-based ranking, we first determine which two successive supermarkets have the largest difference in final score. Next, we determine which two successive supermarkets are separated from each other by the second largest difference in final score. These two large score gaps create three groups of supermarkets: the frontrunners (supermarkets with the highest scores), the laggards (supermarkets with the lowest scores) and the average performers (those in between).

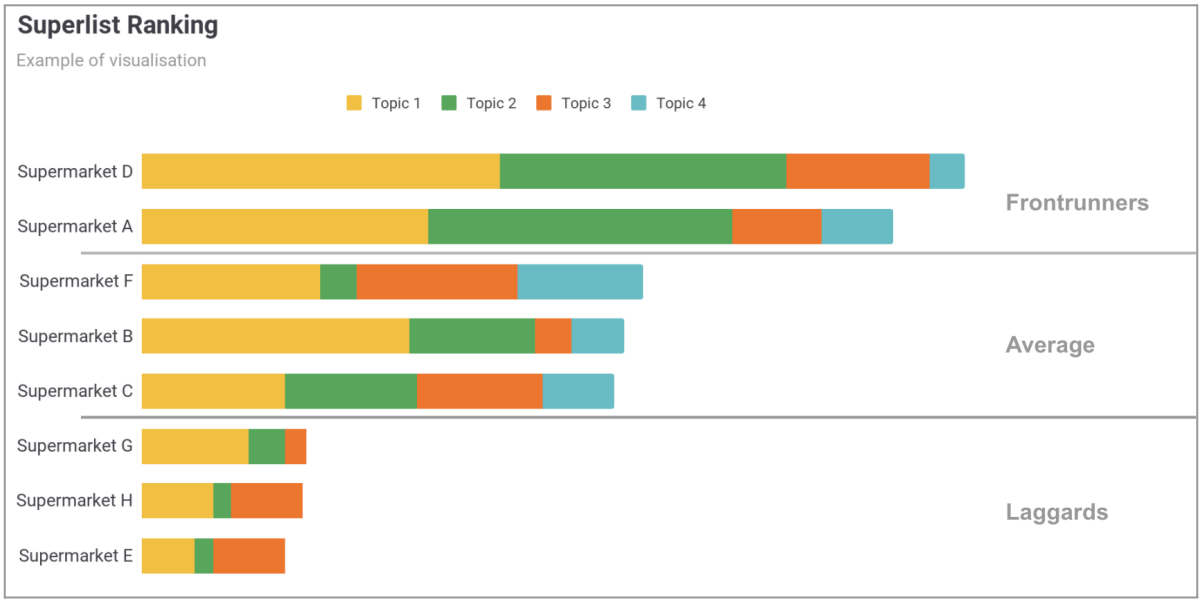

Figure 2. Example: schematic sketch of visual presentation.

The grey lines mark the largest difference in final scores and show where each group – frontrunners, average performers and laggards – begins and ends. (In practice, groups may of course be smaller or larger).

Pilot studies

In certain cases we conduct a pilot study before setting up a regular, full scope Superlist in a certain region. Compared to a regular Superlist, a pilot study is limited in scope, regarding the number of indicators and of supermarkets to be assessed.

The objectives of a pilot study are twofold:

- Spark the public debate around one or two specific health or sustainability issues and the role of supermarkets in addressing them.

- Test the waters for a full scope superlist: assess the need for a benchmark of supermarkets, map the civil society landscape, assess the local public debate around food, find possible funders, etc.

The research methodology for a pilot study deviates in certain respects from this research framework. The most important deviations being:

- For a pilot study we do inform the supermarkets in scope about the upcoming research and the publication, however we do not engage them in the consultation rounds that we typically do for a regular Superlist.

- The indicators in a pilot study do not aim to give a comprehensive picture of issues in the food system, nor do they necessarily relate to the most urgent issues at that moment. Indicators may also be selected for their relevance in the public debate or their alignment with other local initiatives.

- We do not summarise the findings of a pilot study in a ranking of supermarkets. Ranking supermarkets could suggest a comprehensive assessment.

Project Cycle

For each theme, we use the following process to compare supermarkets. This table simply serves as a guideline; a specific planning will be made at the start of each project.