Designing Human Intervention Studies for Scientific Substantiation of Health Claims – How EFSA Thinks

*Steffen Jakobs and Andreas Hahn**

According to Regulation (EC) 1924/2006, health claims made on foods need an authorisation and the effect must be 'established by generally accepted scientific data'. The scientific assessment on applications for health claims is the responsibility of the European Food Safety Authority (EFSA). Thus, the Authority plays the crucial role in the scientific acceptance or rejection of health claims. In accordance with Regulation (EC) 353/2008, in order to substantiate a health claim, data from human studies are required to prove the relationship between the consumption of the food/food constituent and the claimed effect. In this context, EFSA considers double-blind randomised controlled trials in humans to be the gold standard. So far, the Authority has not published an exclusive guidance for reporting of human intervention studies in order to present transparent and consistent criteria for these studies. Thus, clear and concrete guidelines for adequately performing such studies are missing, which represents a general problem for applicants. Therefore, based on an evaluation of all scientific opinions on claims according to Article 13, this paper addresses key factors that EFSA requests to report on the design, conduct and statistical analysis of human intervention studies with foods or food constituents for scientific substantiation of health claims. It also shows that EFSA's work often is neither transparent nor consistent.

Keywords: Health claims; EFSA; Food labelling; Human intervention studies; RCTs.

I. Introduction

1. Background and Objectives

The use of nutrition and health claims in the labelling and advertising of foods in the European Union is regulated by the Nutrition and Health Claims Regulation (NHCR)[1] and has been since 1 July 2007. According to Article 6 para 1 NHCR, 'health claims shall be based on and substantiated by generally accepted scientific data,' and pursuant to recital 23, 'health claims should only be authorised [...] after a scientific assessment of the highest possible standard.' And 'to ensure harmonised scientific assessment of these claims, the European Food Safety Authority [EFSA] should carry out such assessments.' Articles 15 to 18 lay down the procedure for the authorisation of health claims in the individual procedure, whereby, according to Article 16 para. 3 NHCR, ' [...] in preparation for its opinion, EFSA shall review whether (a) the health claim is supported by scientific evidence and (b) the wording of the health claim complies with the criteria laid down in this Regulation [...].' Article 5 Regulation (EC) 353/2008 states that 'studies and other material [...] shall consist primarily of studies in humans' and 'shall be presented according to a hierarchy of study designs, reflecting the relative strength of evidence which may be obtained from different types of studies.' Although Article 5 refers to applications according to Article 15 NHCR, it is ob-

* Dr Steffen Jakobs and Prof Dr Andreas Hahn, Leibniz University Hannover, Institute of Food Science and Human Nutrition, Am Kleinen Felde 30, D-30167 Hannover, Germany. For correspondence: <jakobs@nutrition.uni-hannover.de>; <hahn@nutrition.uni-hannover.de>.

1 Regulation (EC) 1924/2006 of the European Parliament and of the Council of 20 December 2006 on nutrition and health claims made on foods, 2006.

vious that EFSA applies these rules from the beginning. This is in line with the generally accepted scientific approach and cannot be criticised.

Authorised and non-authorised health claims are listed in the EU Register of nutrition and health claims made on foods.[2] There, in addition to other information, a reference to respective EFSA opinions can also be found for every evaluated food-health relationship except for those that had been judged by EFSA but are still 'on hold' (i.e. caffeine). In these opinions, most health claims have been scientifically rejected. Almost 90% of all health claims have proven untenable.[3] This is due to three main reasons: a) The food or food constituent was not sufficiently characterised, b) the postulated health relationship is not beneficial for health (i.e. it is unspecific or not generally advantageous), and c) a cause effect relationship could not be established on the basis of the available data. One main reason was the flawed design and quality of many of the provided studies,[4] particularly human intervention studies. EFSA pointed out lots of inadequacies of these studies in the individual scientific opinions.

In contrast to standard practice,[5] EFSA has not yet published guidance that exclusively addresses requirements that need to be considered when planning, conducting and reporting a human intervention study with a food or a food constituent for a health claim application.[6] Instead, EFSA carries out an individual assessment of study reports or scientific papers that have been submitted as part of health claim applications. The results of these assessments have been stated in many individual EFSA opinions on the scientific substantiation of a health claim.

There is a huge risk for every health claim application since EFSA's scientific judgment often remains unclear and is quite rightly the subject of criticism.[7] The need to define and specify the concept of 'generally accepted scientific data/evidence' in EFSA's administrative practice has already been mentioned.[8] Thus, a deeper understanding of how EFSA thinks and what EFSA demands has become an essential prerequisite for succesful applications. Based on a comprehensive and systematic analysis of all authorised and non-authorised health claims relating to Articles 13.1 and 13.5 NHCR listed in the EU register so far[9] (in total of 484 EFSA scientific opinions with 2057 health claims[10]), this article summarises key factors that address EFSA's requirements regarding the reporting of the design, conduct and statistical analysis of human intervention studies with foods or food constituents for substantiation of health claims – with the intention to provide a practical tool that can be used by applicants for preparing health claim applications.

The results of this article are intended to contribute to what exactly EFSA means by 'well-designed and conducted randomised controlled trials'[11] and to clarify the undefined term of 'generally accepted scientific data' (on which authorised health claims must be based pursuant to Article 6, para. 1 NHCR).[12]

2 See <http://ec.europa.eu/food/safety/labelling_nutrition/claims/register/public/?event=register.home>.

3 Status: August 2019.

4 See: Meyer in Meyer/Streinz, LFGB - BasisVO, commentary, marginal note 87, 2nd edition 2012.

5 EFSA had published several guidance for assisting applicants in the preparation of applications for authorisation of health claims, e.g. General scientific guidance for stakeholders on health claim applications, <https://www.efsa.europa.eu/de/efsajournal/pub/4367>, or Scientific and technical guidance for the preparation and presentation of a health claim application (Revision 2), <https://www.efsa.europa.eu/en/efsajournal/pub/4680>.

6 It must be taken into account that some details on the design, conducting and reporting of human intervention studies are provided in several EFSA guidelines (e.g. Scientific and technical guidance for the preparation and presentation of a health claim application (Revision 2), supra, note 5, pp. 27 et seq.) as well as in the Commission regulation (EC) 353/2008 of 18 April 2008 establishing implementing rules for applications for authorisation of health claims as provided for in Article 15 of Regulation (EC) 1924/2006. Additionally, EFSA held a technical meeting on the reporting of human studies submitted for the scientific substantiation of health claims in 2013, <http://www.efsa.europa.eu/de/events/event/131120#documents>.

7 E.g. Lenssen/Bast/de Boer, 'Clarifying the health claim assessment procedure of EFSA will benefit functional food innovation', Journal of Functional Foods 2018, 47, pp. 386 et sqq.; Thompson/Heneghan/Cohen, 'How valid is the European Food Safety Authority's assessment of sports drinks?', British Medical Journal 2012, pp. 1 et seq.; Hahn/Hagenmeyer, 'EFSA's 'Secret' Health Claims', EFFL 1/2013, pp. 10-24.

8 E.g. Hahn/Teufer, 'Promised is promised! For the scientific substantiation of effect statements about foods with special consideration of human intervention studies', ZLR 2008, pp. 663 et seq.; Biesalski/Aggett/Anton et al., '26th Hohenheim Consensus Conference, September 11 2010, Scientific substantiation of health claims: evidence-based nutrition.' Nutrition 2011 Oct;27(10 Suppl): S1-20.

9 Status: August 2019.

10 Various EFSA opinions, which mentioned an identical relationship between the consumption of a food/constituent and a claimed health effect and for which EFSA did not reassess this relationship, were included as a single opinion in the data analysis.

11 See, for example: Scientific and technical guidance for the preparation and presentation of a health claim application (Revision 2), supra, note 5, p.10.

12 In addition to the requirements regarding the design, conducting and reporting of human intervention studies, the type and scope of scientific references are decisive for a positive decision by EFSA.

Whether and in which cases double-blind randomised controlled trials are necessary is controversially debated[13] and will not be discussed here, nor will the question of how many studies and other literature references are sufficient for EFSA. This article refers solely to these studies because they are usually requested by EFSA.

2. Definition of Health Claims

As defined by Article 2.2 NHCR a health claim means a claim (inclusive of pictorial, graphic or symbolic representation) that states, suggests or implies a relationship between a food, food constituent or food category and health. Also, general, non-specific claims like 'good for your health', 'good for your heart' or 'improve your body defense' have been taken into account by the NHCR (Article 10.3), but an authorisation and a scientific assessment is neither needed nor possible for such claims.

Additionally, the regulation distinguishes between (i) Article 13.1 health claims, (ii) Article 13.5 health claims, and (iii) Article 14 health claims. The latter refers to the reduction of risk of disease or to the growth and development of children, e.g. 'Barley beta-glucans have been shown to lower/reduce blood cholesterol' or 'Docosahexaenoic acid (DHA) maternal intake contributes to the normal brain development of the foetus and breastfed infants.' Article 13.1 health claims (so-called general function claims[14]) are health claims describing or referring to:

1. Role of a nutrient or other substance in growth, development and the functions of the body; or
2. Psychological and behavioural functions; or
3. Slimming or weight control or a reduction in the sense of hunger or an increase in the sense of satiety or to the reduction of the available energy from the diet.

Examples hereof include, '[b]etaine contributes to normal homocysteine metabolism' or '[b]iotin contributes to normal energy-yielding metabolism.'

Health claims pursuant to Article 13.5 rely on newly developed scientific data. Although the NHCR does not clearly state what is meant by 'newly developed scientific data' it can be assumed that a claim based on new scientific evidence is a claim that has not been assessed before by EFSA or is a claim based on evidence that has emerged since 31 January 2008.[15] This deadline had been met by the Member States to submit national lists with proposed health claims, as required by Article 13.2 NHCR. For example, the claim, '[d]aily creatine consumption can enhance the effect of resistance training on muscle strength in adults over the age of 55,' was authorised in April 2017.[16] For a short overview of the different types of health claims on foods, see Figure 1.[17]

II. Methods

This article analyses 484 EFSA opinions on health claims related to Articles 13.1 and 13.5 NHCR. All scientific opinions were searched for specified keywords (see Appendix, Table 1). These keywords have been identified after thorough and multiple reading of the EFSA opinions and exclusively refer to design, conducting, statistical analysis and reporting of human intervention studies with foods or food constituents. The following study designs were considered: single-arm studies; cross-over studies and parallel studies.

The keyword research has been performed by using the lexical search function of the text analysis software MAXQDA (version 10). All hits found in the EFSA opinions have been proven to be relevant to factors that need to be considered when designing, conducting, analysing and reporting human intervention studies to evaluate the health benefits of foods. In most cases, relevant statements about these factors have been found under item 3, 'scientific substantiation of the claimed effect,' and particularly within the sections where the corresponding studies have individually been assessed by EFSA. Based on the identified statements and under consideration of

13 E.g. Biesalski/Aggett/Anton et al., supra, note 8, S15 et seq.; Blumberg/Heaney/Huncharek et al., 'Evidence-based criteria in the nutritional context, Nutrition Reviews 68(8), 2010, pp. 478 et seq.; Ströhle/Hahn, 'In Search of the Evidence-Based Grail - Nutritional Statements in the Era of Evidence-Based Medicine', Aktuelle Ernährungsmedizin 35, 2010, pp. 324 et seq.

14 See <http://www.efsa.europa.eu/en/topics/topic/general-function-health-claims-under-article-13>.

15 See <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/204320/Nutrition_and_health_claims_guidance_November_2011.pdf>.

16 See <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0672&from=EN>.

17 Figure modified from Verhagen/Vos/Francl et al., 'Status of nutrition and health claims in Europe', Arch Biochem Biophys 2010, 501(1):6–15.

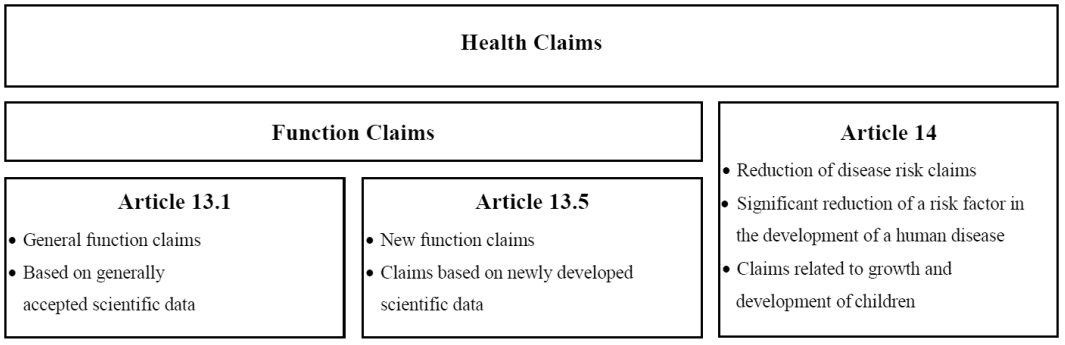

Figure 1: Overview of the different types of health claims corresponding to the NHCR

contextual aspects, conclusions have been drawn for EFSA requirements for human intervention studies with food or food constituents. In many cases, the same keyword has been identified in different scientific opinions. For the sake of simplicity, regularly one or two scientific opinions are identified as a reference for the same hits in different opinions in the results section of the present article. In addition, the highlighted reporting criteria for human intervention studies are summarised as graphical figures (see Appendix, Figures 2 and 3).

III. Results

Conducting human intervention studies is a complex issue that comprises a lot of factors that have to be considered. This refers to the study design itself such as target group, outcome parameter, methods applied and intervention time as well as to the provision of the study and the statistical procedure. Among these, the following aspects are subject of EFSA's criticism.

1. Randomisation and Blinding

In the context of scientific scrutiny of human intervention studies for authorisation of health claims, EFSA examined if randomisation had been carried out. When no or limited information was given about randomisation in the study report or scientific paper, EFSA classified this as a lack of methodical insight.[18] Therefore, the Authority regularly considered that no conclusions for the scientific substantiation of a claim could be drawn from such studies, in particular, if further weaknesses in study design and/or reporting have been noted.[19] A complete scientific assessment of human intervention studies by EFSA needs sufficient information about the randomisation.[20]

Additionally, it should be reported if a blinding assessment has been used, as otherwise non-reporting will be criticised by EFSA.[21] If studies have been conducted without blinding (so called unblinded or open label studies), no conclusions have been drawn from these studies for the scientific substantiation of the claimed effect.[22] Particularly for self-reported results, open label studies have a high risk of bias.[23] EFSA regularly accepted double-blind studies without any complaints respectively blinding and assessed such studies individually.[24] If studies were single-blinded (e.g. nurses and investigators were not blinded to treatments) and EFSA noted additional study flaws, the interpretation of the study results was hampered.[25]

18 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2469>.

19 E.g. <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/j.efsa.2011.2302>.

20 E.g. <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/j.efsa.2014.3656>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2212>.

21 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2012.2888>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2041>.

22 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2254>.

23 See, e.g. Rosenman/Tennekoon/Hill: 'Measuring bias in self-reported data.' International journal of behavioural & healthcare research 2011, 2 (4), pp. 320–332.

24 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1739>.

25 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2380>.

2. Recruitment, Characteristics and Compliance of Study Participants

Human intervention studies should be conducted in a study group that is representative of the population group for which the claim is intended.[26] EFSA had examined separately whether the results obtained in the studied population can be extrapolated to the target population. This was particularly crucial for study results gained in populations that consisted of patients. For instance, osteoarthritis patients were not considered to be representative of the general population regarding the condition of their joint tissues[27] and patients with peripheral arterial disease (e.g. intermittent claudication) were not suitable for investigating the influence of fatty acids on peripheral blood circulation.[28]

Furthermore, EFSA demanded information about the process of recruitment and number of participants recruited. These should be documented in the study report and/or be mentioned in the publication.[29] The Authority also examined whether relevant characteristics of study participants, e.g. age, sex or body weight, are comparable for intervention and placebo group at the beginning of a study (baseline characteristics).[30] Depending on the study hypothesis the following information was usually necessary:

- intake of drugs that potentially influence the study's outcome(s)[31]
- nutritional status[32]
- eating habits/background diet[33]
- physical activity[34] and
- smoking habits[35]

Participants should be selected on the basis of defined and specified inclusion and exclusion criteria.[36] EFSA examined whether and how many randomised participants met the inclusion criteria at baseline.[37] If it was not clear, whether participants' baseline values of outcomes were comparable, no conclusions were to be drawn from these studies.[38] It was also important to provide information about the compliance of study participants[39] and the methods used for assessing them,[40] e.g. capturing the amount and frequency of foods via dietary records/questionnaires,[41] counting of capsules at every study visit[42] or measuring of renal or faecal excretion of consumed food constituents or their metabolites.[43] Partly, quantitative statements regarding the grade of compliance were provided,[44] e.g. 96%[45] or 80%[46] of participants fulfilled the requirements of consumption of the test product. In practice, mostly just qualitative information is provided which does not allow an exact evaluation of the subjects' compliance. Therefore, it is recom-

26 See <https://www.efsa.europa.eu/de/efsajournal/pub/4367>.

27 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2291>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2017.4774>.

28 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1477>.

29 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2012.2996>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2016.4550>.

30 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2057>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1739>.

31 E.g. <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/j.efsa.2010.1730>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2012.3003>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2028>.

32 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2009.1229>.

33 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2009.1242>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2012.2819>.

34 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2057>.

35 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2008.789>.

36 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2017.4839>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2057>.

37 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2014.3842>.

38 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2264>.

39 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2256>.

40 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2057>.

41 E.g. <https://www.efsa.europa.eu/de/efsajournal/pub/2469>.

42 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2302>.

43 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2014.3842>;<https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2033>.

44 See <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1794>;<https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2009.1227>.

45 See <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2074>.

46 See <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2014.3753>.

mended to give a quantitative indication of this criterion.[47]

The drop-out rate of a study was also considered by EFSA and should be included in the study reports or publications.[48] It was noted by EFSA if no information[49] were available or if the drop-outs occurred only in the placebo group.[50] Therefore, the drop-out rate should be reported separately for both the placebo and the intervention group. A high drop-out rate reduces the number of subjects available for statistical analyses, which could limit the reliability of study results and also the conclusions to be drawn therefrom. The following total drop-out rates were criticised as too high: 45%,[51] 40%,[52] 34%,[53] 33%,[54] 31%[55] and 26%.[56] By contrast, an overall drop-out rate of ≤ 20% seems acceptable to EFSA.[57] Furthermore, the reasons for drop-outs should be listed in the study report or the publication.[58] In this context, it is essential to recognise whether drop-outs were systematically caused or not.

3. Study Size and Statistical Power

It was necessary to carry out a sample size and statistical power calculation,[59] which must be based on the primary endpoint of the relevant study.[60] Importantly, all elements of a statistical power calculation must be considered, such as the inclusion of an estimated drop-out rate.[61]

Generally, studies had to be sufficiently large (and have enough statistical power) to detect a statistically significant change in the physiological effect of interest in the intervention vs. control group. EFSA classified a small sample size of a study as a methodological weakness[62] and when studies were of adequate respectively large size; this was highlighted by EFSA in some cases.[63] It must be emphasized that statistical significance is only one prerequisite. The outcome effect must be biological and/or clinically relevant, too. This should be verified by clinical trial investigators but in most cases, it remains unclear which effect size will be classified as 'relevant' by EFSA. For example, to what extent must blood flow, cholesterol levels or postprandial blood glucose levels be influenced for EFSA to regard the effects as biologically relevant? Indeed, only a handful of EFSA opinions address the issue of the biological/clinical relevance of the study results.[64] Therefore, the assessment of clinical/biological relevance should be included as standard point in EFSA's scientific opinions in the future.

47 See Welch/Antoine/Berta et al., 'Beyond Passclaim – Guidance to substantiate health claims on Foods – Guidelines for the Design, Conduct and Reporting of Human Intervention Studies to Evaluate the Health Benefits of Foods' Br J Nutr 2011, 106: Suppl. 2: S3–S15.

48 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2013.3083>.

49 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2009.1235>.

50 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2227>.

51 See <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1725>.

52 See <https://www.efsa.europa.eu/de/efsajournal/pub/2469>.

53 See <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2012.2949>.

54 See <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1798>.

55 See <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1860>.

56 See <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2264>.

57 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2210>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2264>. Even if the reporting of an acceptable drop-out rate depends on various factors such as survey mode, study design and research question and needs to be considered on a case-by-case basis, it would be desirable that EFSA would establish an acceptable drop-out rate for guidance.

58 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2266>; <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/j.efsa.2011.2302>.

59 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2008.789>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2212>.

60 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2009.1259>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2263>; <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/j.efsa.2011.2023>.

61 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1739>.

62 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1773>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2047>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2057>.

63 For example, in the opinion on olive biophenols under point 3.1 in the multi-center study by Covas et al. (2006b) with a total of 200 study participants; see <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2033>.

64 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2008.853>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2008.789>.

4. Control Group and Placebo

In general, EFSA scientific opinions mentioned whether studies were performed with a control group or not. Uncontrolled studies (single arm studies) were classified as non-pertinent and therefore excluded from scientific evaluation, in most cases by EFSA[65] and in rare cases by applicants[66] in the context of the identification of pertinent studies. Nevertheless, according to the Annex of Regulation (EC) 353/2002, the totality of the available scientific data should be taken into account for the substantiation of health claims. However, uncontrolled studies alone are not sufficient for a positive EFSA decision.

EFSA demanded that the substance (or substances/food constituent) used as a placebo is clearly identified and specified. It was important to state the exact composition of the placebo in the study report or publication. Ideally, placebo and intervention products (including test and control foods) are comparable in appearance, colour, smell, texture, taste and packaging, to exclude a distinction by such properties by study subjects and to avoid an expectation bias.[67]

For human studies that were conducted with control foods (including fortified foods) EFSA set out requirements for essential characteristics of these foods. In contrast to the test food, the control food should not contain the effect triggering substance (or substances), otherwise, EFSA did not take these studies into account.[68] Test and control foods should be matched for energy content, macronutrient composition (such as carbohydrate, fat, and protein content) and other relevant nutrients.[69] The same applies to specific diet programs used as nutritional intervention and placebo. It must be assured that changes in surrogate parameters or endpoints (e.g. body weight) are clearly attributable to a corresponding food ingredient (e.g. soy protein) and are not caused by different levels of macronutrients in tested study diets.[70] It is obvious that nutrition studies usually cannot be conducted using a 'real placebo' if foods[71] (like fruits or vegetables) or whole diets instead of food constituents are subject of the study. It must be considered that nutrients occur ubiquitously and that a normal diet cannot be free from the food ingredient to be investigated. The conduct of this type of randomised controlled studies (RCT) would also be ethically unacceptable.[72]

5. Dosing and Dose-Response Relationships

EFSA requested to provide information on the (daily) quantity of food or the dosage of a food ingredient[73] (including reporting of colony forming units (cfu) for probiotic bacteria[74]). The Authority usually noted when dose-response relationships[75] or no dose-dependent effects were found in studies. If no information on this issue were available and additional study flaws stated, the corresponding study was at risk of being excluded.[76] Although dose-response relationships are undoubtedly helpful, they are not always present. In no way can they be considered a prerequisite for positive EFSA assessments, since there are also numerous approved health claims for which EFSA did not mention dose-response curves in its opinions.[77]

Studies using a dosage significantly above the Tolerable Upper Intake Level (UL) normally have not

65 There are countless examples of this in the EFSA opinions, e.g. the study by Konrad et al. (1999) in the EFSA opinion on alpha lipoic acid and its possible influence on insulin sensitivity; see <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2202>.

66 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2012.2809>.

67 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1792>.

68 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2211>; a positive example is the study by Shimomura et al. (2006), see <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1790>.

69 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2252>;<https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1792>.

70 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1812>.

71 Except for fortified foods.

72 E.g. Biesalski/Aggett/Anton et al., supra, 8, S15 et seq.; Blumberg/Heaney/Huncharek et al., supra, note 13, p. 480 et seq.

73 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1689>.

74 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2009.999>.

75 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2008.789>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2264>; <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/j.efsa.2011.2040>.

76 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1792>.

77 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2012.2694>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2014.3513>.

been accepted by EFSA.[78] In the context of the evaluation of niacin and the normal maintenance of LDL and total cholesterol levels all submitted studies were excluded from the scientific evaluation because they used a daily niacin dosage of 100 mg to 6 g, which was significantly above the UL of 10 mg niacin per day. It was stated that the mentioned health claim about niacin was contrary to Article 3 c) NHCR, which prohibits that health claims encourage excessive consumption of food.[79] A very high daily dosage does not reflect the actual situation of the usual alimentary intake of a nutrient[80] and the question arises whether these are pharmacologically active dosages.

Overall analyses of EFSA opinions show that designing studies with oral administration of megadoses of nutrients for the substantiation of health claims is superfluous, as these are not regarded as valid by EFSA. In accordance with the requirements of Article 5, para. 1 d) NHCR positive effects should be triggered at doses that can be achieved with a reasonable amount of a food or with a normal diet. However, EFSA also noted when studies used a lower dose as specified in the proposed conditions of use,[81] the dosage in the studies should in each case be consistent with the conditions of use in the application.

6. Study Duration

The duration is an essential characteristic of a trial. The Authority mostly stated in its scientific opinions how long a study was conducted and whether the study duration is suitable to match with the conditions of use of a respective health claim.[82] If studies were too short, this was explicitly mentioned[83] and if additional study flaws were found by EFSA, such as a small number of participants or absence of a control group, these studies were excluded from a scientific assessment.[84]

It is important, at what point in time a change of a relevant effect (or variable) occurs or is observable in a study. Ideally, this can be derived from previously performed (exploratory) studies,[85] mechanism of actions, biochemical or physiological aspects. The length of a trial must, in any case, be sufficient to measure a change in the effect of interest.[86] No general recommendations regarding the determination of the duration of the study can be given, rather this depends on the relevant study question or hypothesis, the food (or food ingredient), the time required to measure the appropriate endpoint and the available financial budget.

Partially EFSA set requirements for study duration in its specific scientific guidelines. For instance, the scientific substantiation of health claims related to the influence of homocysteine concentrations should be based on studies with duration of eight weeks for measuring long-term effects.[87]

7. Validity of Methods and Outcome Variables

Usually, EFSA reviewed which methods were used to measure the relevant outcome(s)[88] and the ingestion of a food or food ingredient.[89] If no information was available, no scientific evaluation of the correspond-

78 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2009.1229>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2008.905>.

79 See <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2009.1224>.

80 Aggett/Antoine/de Vries et al., 'Beyond PASSCLAIM – Guidance to substantiate Health Claims on Foods ', Summary Report of a Workshop held in December 2009 in Nice, France, in: ILSI Europe Report Series, Brüssel 2010.

81 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2008.789>.

82 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1794>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1798>; However, the adequacy of the duration of a study is not stated in every individual assessment of a study.

83 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2014.3756>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2016.4480>.

84 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2051>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2054>

85 E.g. <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/j.efsa.2014.3654>.

86 E.g. EFSA questioned whether the duration of the intervention in the study by Hewlett and Smith (2007) was sufficient to measure the impact of caffeine on a sustained increase in attention. See <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2054>; similar to: <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2229>.

87 See <https://doi.org/10.2903/j.efsa.2018.5136>.

88 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2241>.

89 For example, under laboratory conditions or through self-assessment by the participants; e.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1464>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1761>.

ing studies was carried out.[90] Additionally, it was examined whether endpoints or biomarkers had been determined with suitable and established methods.[91] If non-valid methods or biomarkers were used in studies, no conclusions were drawn from them.[92] For example, EFSA requested proof of validation of questionnaires that were used for data collection.[93] Likewise, the assessment of subjective sensations, e.g. quantifying muscle pain with visual analogue scales[94] or evaluating individual IBS symptoms with symptom scores,[95] required validation of the measurement scales. Particularly in the case of self-assessed outcomes EFSA requested that methods were validated.[96] In this context, the Authority criticised if a self-assessment of outcomes intended for the beginning and end of a study was only carried out at the end of the study, since the likelihood of a recall bias rises.[97]

The specificity of a measurement method was also assessed by EFSA.[98] For instance, the Authority noted that a measure of trans-epidermal water loss (TEWL) is a direct outcome measure that can be used for the scientific substantiation of a health claim on protection of the skin against dehydration. When an outcome, such as intestinal transit time, was measured by a variety of methods, but only some measurements showed a statistically significant change in outcome, EFSA categorised these findings as inconsistent.[99]

It was also considered whether the claimed effect of a food or food constituent was measured as the primary endpoint.[100] If the relevant outcome was measured as secondary endpoint, EFSA normally stated that the study results '[...] for changes in [an outcome variable] as secondary outcome are at increased risk of bias.'[101] So, no conclusions were drawn from these studies if additional shortcomings had been present. However, the Authority did not proceed consistently in this matter. For example, EFSA's positive opinion on native chicory inulin and increasing stool frequency is based on a number of studies in which EFSA even notes that 'stool frequency was only a secondary outcome in the majority of the studies'.[102] Nevertheless, EFSA issued a positive decision in this case. In contrast to other opinions,[103] EFSA did not mention in this case that a correction for multiple testing in relation to the statistical analysis of several secondary endpoints is necessary. If several primary endpoints were identified, EFSA requested that the multiplicity of endpoints must be adequately reflected in the statistical analysis (see section III.8.c).[104]

8. Statistical Analysis

a. Per-Protocol and Intention-to-Treat Analysis

EFSA explicitly requests a proper statistical analysis of studies in line with its design and a sufficient description of the statistical methods.[105] Accordingly, the Authority criticised if no information about the statistical methods was given in a study protocol or publication.[106] In addition, it was demanded that da-

90 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2265>.

91 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2015.4095>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1794>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2009.1271>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2033>.

92 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2033>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1753>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2031>.

93 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2009.1238>.

94 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1753>;<https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1804>.

95 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2008.853>.

96 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1465>;<https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2009.1235>.

97 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2037>.

98 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1463>.

99 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1817>.

100 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2224>.

101 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2469>.

102 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2015.3951>.

103 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2469>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2017.4723>.

104 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2469>; EFSA-Gutachten Nr. 257, 2011, S. 6; EFSA-Gutachten Nr. 311, 2011, S. 6ff.

105 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2264>.

106 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2212>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2264>.

ta on the total number of subjects included in the statistical analysis have been provided[107] and that the analysis of data did take into account baseline values.[108] EFSA also checked whether intergroup comparisons of study outcomes had been performed[109] and whether group differences were statistically significant.[110] When only intragroup comparisons were carried out, EFSA excluded such a study from further consideration.[111]

A crucial question was how protocol violations of study participants were managed.[112] For this purpose, studies were analysed either according to per-protocol analysis (PP analysis) or intention-to-treat analysis (ITT analysis). For some trials, the results of both methods were reported.[113] A PP analysis is a method that excludes all subjects or patients from the statistical analysis who did not adequately adhere to the protocol (e.g. those who did not take all the intended treatment, received a different treatment or no intervention, or terminated prematurely from the study). When only a PP analysis was carried out, EFSA stated that this would severely limit the conclusions to be drawn from the study.[114]

An ITT analysis includes all randomised participants and retains all of them in the group to which they were allocated ('once randomised always analysed'). So, the data analysis is performed independently of whether participants changed the study group, whether there was insufficient subject compliance or whether the study was discontinued. A PP analysis is potentially more susceptible to biasing results, whereas an ITT analysis describes an effect under real world conditions.[115] EFSA highlighted when no ITT analysis was carried out[116] and considered non-application of ITT analysis to be a methodological deficiency.[117] EFSA also requested evidence to justify the discrepancy between results of the various analyses ((such as PP and ITT analyses, if applicable).[118] The Authority further requested that longitudinal studies be analysed using appropriate statistical methods, e.g. repeated measures ANOVA (RM-ANOVA) or mixed models.[119]

b. Subgroup Analysis

If subgroup analyses were carried out on patients who suffered from a disease with different subtypes, EFSA requested prior stratification and thus the formation of corresponding subgroups. For instance, when assessing a study on the claimed relationship between *Lactobacillus paracasei* B21060a and a reduction in gastrointestinal discomfort, EFSA criticised a subgroup analysis of 47 patients with irritable bowel syndrome (IBS) of the diarrhea-predominant type for not performing a priori stratification in the various subtypes of IBS (diarrhea-predominant type, constipation-predominant type and mixed type).[120] For this reason, EFSA concluded to be unable to draw any conclusions from this subgroup analysis. If randomisation did not consider any planned subgroup analyses and further study deficiencies were present, EFSA excluded the corresponding subgroup analysis.[121] Thus, stratified randomisation should be considered for the proper performance of subgroup analysis to prevent

107 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2270>.

108 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2012.2888>.

109 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2014.3753>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2229>.

110 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2208>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1798>.

111 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2231>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2264>.

112 Welch/Antoine/Berta et al., supra, 47, S13.

113 E.g. <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/j.efsa.2011.2302>; in a few cases, EFSA also stated that analyses were carried out as modified intention-to treat analysis (mITT); see, e.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2016.4539>.

114 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1725>.

115 See Ranganathan/Pramesh/Aggarwal, 'Common pitfalls in statistical analysis: Intention-to-treat versus per-protocol analysis' Perspect Clin Res. 2016; 7(3), pp. 144–146.<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4936074/>; Welch/Antoine/Berta et al., supra, note 47, S13.

116 E.g. <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/j.efsa.2011.2023>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1768>.

117 E.g. <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/j.efsa.2009.1267>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2012.2949>

118 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2014.3753>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2016.4538>.

119 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2013.3083>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2016.4539>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2016.4455>.

120 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1804>.

121 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2264>.

randomisation bias. If subgroup analyses were not planned in advance (pre-specified) but carried out post-hoc, EFSA did not draw any conclusions from these subgroup analyses or stated that this methodical weakness greatly limits the value of a study.[122] For post-hoc subgroup analyses it was necessary to run a test of interaction effects. Without such testing, the subgroup analysis was excluded from consideration when further flaws occurred (e.g. no reporting of the number of subjects in the individual subgroups).[123]

c. Missing Values, Confounders and Multiple Testing

EFSA also examined how the missing values of subjects' outcomes (e.g. due to drop-outs) were handled in statistical analyses. The problem of missing values is of considerable relevance regarding potentially biased results and a reduced validity of study results.[124] If missing values were not included in the statistical analysis at all, EFSA noted this in the individual assessment of a study.[125] Basically, there are different imputation methods for missing values available. One of these is the Last Observation Carried Forward (LOCF) method, which uses the last recorded value of a participant, who dropped out, as an estimator for all unknown values of a subject until the study was finished. However, EFSA has repeatedly criticised the use of the LOCF method and has not considered it in most cases as a suitable imputation technique for missing values.[126] In three opinions,[127] EFSA deviates from its negative stance and accepts (without critical comment) statistical analyses using the LOCF method.[128] This once again illustrates EFSA's inconsistent approach. What imputation method is sufficient (from EFSA's point of view) cannot be clearly derived from the analysis of the EFSA opinions.[129] This is particularly controversial because EFSA does not accept the subsequent replacing of participants who dropped out during a study after randomisation.[130]

The Authority complained when, during a study, other influencing factors besides the actual intervention (confounders) possibly occurred, such as medication[131] or the alimentary uptake of other nutrients,[132] affecting the outcome with no appropriate control to counter it.[133] Therefore, it was checked whether an adjustment for possible confounders was made.[134] EFSA also noted whether the adjustment for potential confounders such as age, gender or weight had changed corresponding results (or the statistical significance of results).[135]

For randomised controlled trials with a cross-over design and without a washout period, EFSA assessed whether timing and sequence adjustments were made to exclude carry-over effects.[136] If multiple end-

122 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1422>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2016.4366>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2264>.

123 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2009.1259>.

124 See Mayer, 'Dealing with missing values in longitudinal studies in the case of drop-outs', in Hilbert/Minkenberg (ed.), Proceedings of the 16th Conference of SAS users in research and development (KSFE), 2012, pp. 207 et sqq., at p. 208.

125 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2017.4914>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2264>.

126 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2469>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2264>.

127 To date.

128 See <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2016.4455>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2016.4538>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2012.2715>.

129 E.g. <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/j.efsa.2011.2291>; In one opinion, EFSA stated that the application of a linear multi-level model was considered suitable for handling missing values. See <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2016.4455>. However, mixed linear models are not a technique for data-imputation per se, yet merely constitute a general framework for analysing nested or clustered data (i.e. observations within probands; probands within study-centres, etc.). Although mixed linear models allow for the imputation of data, EFSA should specify the imputational technique that was used in the corresponding study (e.g. full information maximum likelihood (FIML) or restricted maximum likelihood (REML)).

130 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2264>.

131 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2014.3753>.

132 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2012.2602>.

133 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2009.1256>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2050>.

134 See <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2241>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2008.853>.

135 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2015.4097>; <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/j.efsa.2011.2302>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1469>.

136 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2264>; <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/j.efsa.2016.4480>.

points (or hypotheses) have been statistically tested in a study, EFSA demanded that it be adjusted accordingly and that a correction for the level of significance is needed.[137] For this purpose, several methods are available, such as the correction corresponding to Bonferroni,[138] Bonferroni-Holm[139] or Benjamini-Hochberg,[140] which have been accepted by EFSA. Furthermore, multiple testing poses the risk that differences between intervention and placebo groups are coincidental, in particular if only borderline statistical significance has been detected.[141] However, all highlighted reporting criteria for human intervention studies are summarised in Figures 2 and 3 (see Appendix).

IV. Discussion

European NHCR was established in order to avoid misleading consumers on nutrition and health claims. Nutrition claims can be made if the respective claim is listed in the Annex and all other requirements of the regulation are fulfilled, especially those resulting from Articles 5 and 6. In contrast, health claims must be authorised 'after a scientific assessment of the highest possible standard' as recital 23 states. This task was transferred to EFSA without giving concrete information to EFSA as well as to applicants on how to deal with these requirements. Thus, EFSA writes the rules for judging the science behind health claims. This procedure raised criticism for several reasons as neither transparent nor consistent.[142]

In contrast to the evidence grading systems of the World Cancer Research Fund (modified by the WHO)[143] and of Richardson respectively of the Passclaim Project,[144] which categorises the evidence respectively according to their respective strength (convincing, probable, possible and insufficient evidence), European NHCR does not differentiate between evidence strengths and demands the highest possible standards in scientific assessment. Thus, it is not surprising that EFSA developed a perspective on claims applications that established a high standard. From a scientific point of view this approach is appropriate because it reflects the generally accepted rules in science. But it must be doubted that this is suitable with regard to the regulatory purpose of NHCR since it implies that all claims not yet finally proven are classified as misleading without any differentiation. The consequences are well known including the high number of rejections as well as the fact that practically no claims on botanicals will presumably be authorised.

Up to now it cannot be expected that the European regulator is going to ameliorate its own regulation and to implement changes that would be necessary for proportionality of the regulation. For that reason, the present article tried to gain deeper insight into how EFSA thinks when evaluating claims applications by analysing 484 scientific opinions according to Articles 13.1 and 13.5 NHCR, which comprise 2.057 claim assessments. There is no need to emphasise that EFSA generally demands for human studies, in most cases intervention studies, to determine whether a cause and effect relationship between a food or a food ingredient and a health effect should be established.

The reason for EFSA opinions with a negative decision in the third assessment step (scientific substantiation) very often referred to the study design, the methodological approach or the statistical analysis of human intervention studies. The in-depth analysis shows (Figures 2 and 3 in Appendix) that a number of criteria that EFSA uses can be identified. It should be noted in particular that the EFSA also called for improved reporting on these aspects of the study. Improved reporting can increase both the va-

137 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2016.4539>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2014.3756>; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2263>.

138 See <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2258>.

139 See <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2291>.

140 See <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2016.4539>.

141 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2231>.

142 E.g. Lenssen/Bast/de Boer, 'Clarifying the health claim assessment procedure of EFSA will benefit functional food innovation', Journal of Functional Foods 2018, 47, pp. 386 et sqq.; Thompson/Heneghan/Cohen, 'How valid is the European Food Safety Authority's assessment of sports drinks?', British Medical Journal 2012, pp. 1 et seq.; Hahn/Hagenmeyer, 'EFSA's 'Secret' Health Claims', EFFL 1/2013, pp. 10-24.

143 See <https://apps.who.int/iris/bitstream/handle/10665/42665/WHO_TRS_916.pdf;jsessionid=27DF67D255782C4E234BCFE138BF89BB?sequence=1>.

144 Richardson/Affertsholt/Asp et al., 'PASSCLAIM, The Process for the Assessment of Scientific Support for Claims on Foods – Synthesis and review of existing processes, European Journal of Nutrition 42, 2003, Supplement 1: I/96–I/111, pp. I/106 et seq.; Richardson: 'The scientific substantiation of health claims with particular reference to the grading of evidence,' European Journal of Nutrition 44, 2005, pp. 319–324; pp. 320 et seq.

lidity of the study results and their transferability to the target group. In addition, the availability of comprehensive study information enables an adequate scientific evaluation of the submitted studies.

EFSA's approach is only partially transparent. So far, no exclusive guidance has been published on reporting studies to assist applicants and on disclosing the requirements of the Authority for the design, conduct and statistical analysis of human intervention studies.[145] Instead, relevant information must be extracted and compiled from a large number of EFSA opinions, which has been attempted in the context of this work. Unfortunately, the explanations in these opinions on what EFSA considers to be 'correct reporting' are very brief and not always clear. For certain aspects they are almost completely missing, as the example of imputation methods (replacement of missing values) shows. EFSA, for example, criticised the use of the LOCF method for replacing missing values as unsuitable, without making clear suggestions or specifications as to which imputation method to use from its point of view.

It is also clear (not described in detail) that EFSA does not proceed consistently. On the one hand, it usually places the highest possible demands on the quality of human intervention studies[146] — which are comparable to those of drug studies — for the scientific substantiation of health claims on non-essential (other) substances (such as creatine[147] or dietary fibres from sugar beets[148]). This is not to be criticised scientifically, but must be viewed critically with regard to the proportionality of the NHCR. On the other hand, the Authority assessed positively a great number of health claims on essential nutrients (such as vitamins and minerals) without the submission of human intervention studies. Examples of this are: Iodine contributes to normal energy-yielding metabolism,[149] folate contributes to the reduction of tiredness and fatigue[150] and thiamine contributes to normal psychological function.[151] In the case of the above examples, it was — quite understandably — sufficient for EFSA to rely on the findings on the biochemical functions of a nutrient. It was clearly concluded from the functional limitations associated with a deficiency that the appropriate nutrient is required for the function ('contributes to normal function of'). Thus, a different assessment standard and requirement level for the scientific substantiation of essential and non-essential substances can be clearly seen here.

The requirements (used by EFSA) for human intervention studies in the context of the authorisation of health claims made on foods are partly found in a similar form in the CONSORT[152] Statement — but not in this specificity — as well as in the publications resulting from the PASSCLAIM[153] project and the workshop Beyond-PASSCLAIM.[154] In principle, this underlines that EFSA's approach could not be criticised scientifically if it were more transparent and consistent. However, it should be noted that the CONSORT Statement, to which EFSA explicitly refers,[155] was mainly designed for medical interventions and also takes into account, for example, their side effects. Safety aspects do not play a role in EFSA's scientific assessment of health claims made on foods.

It should be noted that the publication of a study in a peer-reviewed scientific journal does not guaran-

145 EFSA provides some general information on reporting on certain aspects of study design and statistical analysis. However, these are not specific to health claims on food, the use of concrete statistical methods recommended by EFSA and the way statistical analyses should be carried out. See Guidance on Statistical Reporting; <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2014.3908>.

146 EFSA's approach is in line with recital 23 HCVO, where the legislator states that the use of health claims in the EU should only be authorised following a scientific assessment at the highest possible level. See also Scientific and technical guidance for the preparation and presentation of a health claim application (Revision 2), supra, note 5, at pp. 3 et seq.

147 See <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2016.4400>.

148 See <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2011.2468>.

149 See <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2009.1214>.

150 See <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1760>.

151 See <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1755>.

152 Consolidated Standards for Reporting Trials. See Schulz/Altmann/Moher, 'CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomised Trials', PLOS Medicine 2010, 7(3), pp. 1 et sqq.; EFSA referred explicitly to the Consort Statement within the technical meeting on the reporting of human studies submitted for the scientific substantiation of health claims (2013); see <http://www.efsa.europa.eu/de/events/event/131120#documents>.

153 See, for example, Aggett/Antoine/ASP et al., 'PASSCLAIM, Consensus on Criteria', European Journal of Nutrition 2005, 44, Supplement 1: I/5–I/30.

154 Welch/Antoine/Berta et al., supra, note 47.

155 See <http://www.efsa.europa.eu/de/events/event/131120#documents>.

tee that the study will be considered appropriate by EFSA. In a number of cases, such studies received no or only limited consideration, for example due to an inadequate or missing description of randomisation, blinding or other important aspects of the study. It is obvious that EFSA's reporting requirements for human intervention studies are not necessarily the same as those for peer-reviewed studies. This is also understandable to the extent that the review process is often as heterogeneous as it is non-transparent, so that reviewers sometimes positively evaluate inadequate papers, while others reject appropriate publications with excessive requirements.

The suitability of already published studies for the scientific substantiation of health claims should therefore be examined very carefully in future applications. In particular, the test product used (including its dosage) and the surrogates/endpoints investigated must correspond exactly to the food respective of the claimed effect of the health claim. Given the scope of study data requested by EFSA, it is recommended that the full study report[156] (not only for pertinent unpublished or proprietary data) be submitted,[157] if available. This can help to avoid delays in the processing of applications due subsequent requests, e.g. to carry out additional statistical analyses[158] and to provide additional data[159] under EFSA's 'stop the clock' procedure.[160] This is particularly useful in the light of the fact that 75% of stop the clock events have so far been triggered by clarifications on the studies submitted for scientific substantiation.[161]

When interpreting the results of this article, it should be noted that EFSA's assessments are made on a case-by-case basis and are not always drawn up according to the same pattern and requirement profile. Therefore, the findings presented can only ever be seen as an indication of EFSA's approach. It is recommended to take these into account when submitting an application in order to ensure the quality of human intervention studies required by EFSA. However, it must be taken into account that requirements may change over time as a result of new findings, e.g. in the study methodology or statistical methods, as well as changing views within the NDA panel responsible for scientific assessment of health claims. This panel is newly staffed with members every four years (status at present).[162] The problematic aspect is that there are still no criteria published by EFSA according to which the various scientists can assess health claims uniformly throughout Europe[163] and which could also give applicants more certainty.

V. Conclusion and Outlook

The EFSA requirements for human intervention studies (with parallel group and crossover design) identified in this paper can serve as a guide for future adequate planning of study design and reporting of study methods and results. They enable applicants to review human intervention studies in order to ensure compliance with EFSA requirements and thus prepare their application documents more effectively. However, there is still a lack of exclusive guidance from EFSA that would make its approach transparent. This is to be hoped for as a matter of urgency, especially since a legally binding provision as to which assessment criteria are to be used is not to be expected. In addition, it remains to be seen whether EFSA will impose its own high standards of reporting on human studies on itself (and not only on applicants) and improve its own reporting in its scientific opinions in order to present it more transparently and systematically.

156 For necessary information in the full study report (from EFSA's point of view), see Scientific and technical guidance for the preparation and presentation of a health claim application (Revision 2), supra, note 5, at pp. 27 et seq.

157 E.g. <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2010.1774>.

158 As seen, for example, <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2014.3756>.

159 As seen, for example, <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/j.efsa.2014.3656>.

160 See <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2014.3553>.

161 See General scientific guidance for stakeholders on health claim applications, supra, 5, at pp. 32 et seq.

162 See <https://www.efsa.europa.eu/de/panels/nda>.

163 Hahn/Teufer, supra, note 8, at pp. 681 et seq.

Appendix - Table 1

Table 1: Keywords used for elaborating EFSA requirements regarding the design, conducting, statistical analysis and reporting of human intervention studies with foods or food constituents

Requirements	**Keywords or combined keywords**
adjustment and confounder	corrected, correction, adjustment, adjusted, adjusting, was confounded, confounder, confounding
biomarkers, endpoints, methods of measurement and outcomes	biomarker, marker, measurement, methods, measured, measures, measuring, recall bias, expectation bias, surrogate, self-reported, self-assessed, self-rated, scale, score, item, validation, questionnaire, validated, primary outcome, primary endpoint, secondary endpoint, secondary outcomes
blinding	blinding, blind, double-blind, single-blind, unblinded
carry-over effects	carry-over, carry over, washout period
comparisons within groups and between groups	within-group, within group, between-group, between group, comparisons, compared, intergroup, intragroup
compliance and dropout rate	compliance to, compliance with, compliance of, complied, compliance, non-compliance, compliant, drop outs, drop-outs, drop out, drop-out, dropped out, complete, completed, completers, completion
control group and placebo	control, placebo-control, controlled, uncontrolled, non-controlled, placebo, placebo controlled, placebo-controlled
dose and dose-response relationships	dose-dependent, dose-response, dose, doses, dose-dependency, dose-related, intakes of the food, daily consumption, cfu
missing data/values	missing data, missing values, carried forward, carrying forward observations, last observation carried forward, imputation, impute(d), imputing
multiple testing	multiple testing, multiple comparisons, chance finding, multiplicity, Bonferroni, Bonferroni-Holm, Hochberg correction method
power calculation, sample size and estimation study size	sample size, power calculation, calculated, powered, level of significance
randomisation	randomization, randomisation, randomized, randomised, non-randomised, randomly, random order, order of the intervention, order of the diets, intervention order, allocate, allocation
recruitment and selection of subjects	recruitment, recruitment of subjects, recruited, highly selected sample of subjects, selected
statistical analysis	intention-to-treat, intention to treat, ITT, per protocol, per-protocol, per protocol, completer analysis, (PP) analysis, data-analyses, completed the study, completers, statistical analyses, RM-ANOVA, mixed model

Requirements	**Keywords or combined keywords**
stratification and subgroup analysis	sub-grouping, sub-group, subgroup, subgroups, subgroup analysis, were stratified, stratification, post-hoc, posthoc, pre-planned
study duration	short duration, lasted, time, single occasion, shortterm, short-term, longterm, long-term
participants' characteristics/ inclusion and exclusion criteria	background diet, eating habits, smoking, subjects, physical activity, drug, medications, baseline characteristics, participants, inclusion criteria, exclusion criteria

Appendix - Figure 2, Figure 3

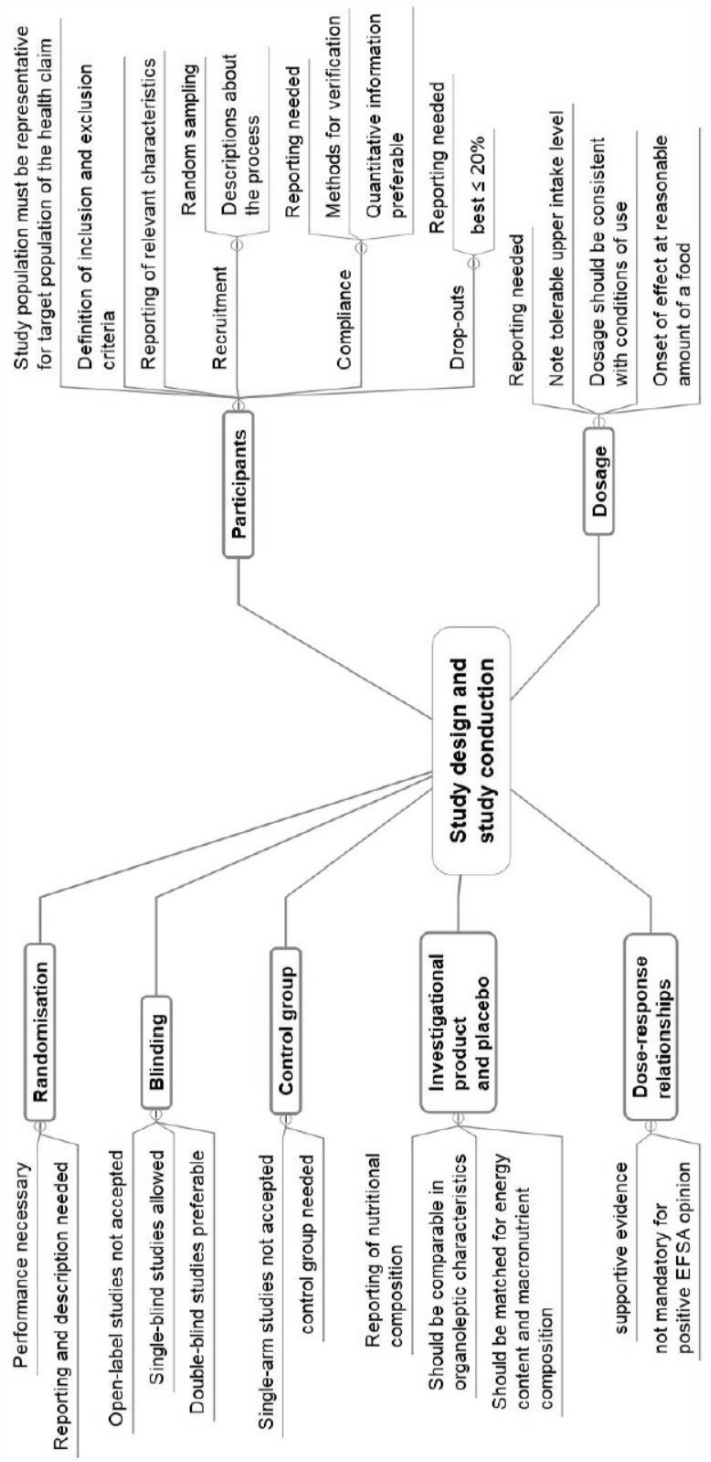

Figure 2: Requirements for reporting of design, conduction or statistical analysis of human intervention studies suitable for health claim substantiation – Part 1 (own figure)

Figure 3: Requirements for reporting of design, conduction or statistical analysis of human intervention studies suitable for health claim substantiation – Part 2 (own figure)