**Paper SI05**

# Let food be thy medicine and medicine be thy food

Fatima Kassim, OCS Life Sciences, 's-Hertogenbosch, The Netherlands
Louella Schoemacher, OCS Life Sciences, 's-Hertogenbosch, The Netherlands

## ABSTRACT

*"Dear diary, this morning I had one bowl of yoghurt. I drank one glass of grapefruit juice and one glass of water for my medication."*

The person who wrote this diary is exposed both to nutrition and pharmaceuticals. However, the way data would be collected and standardised in clinical trials is very different. How do these differences affect data standardisation across these industries? The challenges lie not only in the way data is collected, but there are also stricter laws governing one over the other. The nutrition Therapeutic Area User Guide (TAUG) was published in September 2019, so will this lead to better standardisation practices? In this paper we present viewpoints on this subject and hopefully guide others to a meaningful discussion on the implementation of standards for nutritional data. What is the added value of using CDISC standards for nutritional data and what can be learned from pharma and the other way around?

## INTRODUCTION

The testing of different therapies for the treatment of any ailment or disease is an extremely old practice, with the first one conducted in the modern era to be the Scurvy trial in 1747 by Dr. James Lind. This trial found that 'oranges and lemons' (i.e. Vitamin C) provided the best outcome.[1] The use of clinical trials and their evolution to the standard we know today (i.e. a randomised, double-blind controlled trial) have greatly contributed to the exploration and expansion of clinical care and pharmaceutical research. Clinical trials in the pharmaceutical industry are both similar and different to clinical trials in the nutritional industry. Both industries involve participation of human subjects, resulting in carefully having to adhere to the declaration of Helsinki and Good Clinical Practice guidelines. Also, in both industries the Randomised Controlled Trial (RCT) is considered to be the gold standard. [2]

Modern day nutrition science, unlike chemical or pharmaceutical sciences, finds it roots at the beginning of the 20$^{th}$ century.[3] Nutrition research into complex diseases has only increased in the last couple of decades, especially in the last two, and this leads us to the main issue: what to do with all this data?

This is a question that has plagued the pharmaceutical industry as well. To combat this, harmonisation efforts have been advocated for and successfully developed and implemented by the International Council for Harmonisation (ICH) since 1990 and the standardisation approaches by the Clinical Data Interchange Standards Consortium (CDISC) since its inception in 1997. Thus, the pharmaceutical industry has had a couple of years of experience when dealing with standardisation of complex and big data. Not only has standardisation been guided by CDISC and encouraged by the ICH, it is often encouraged or even required by regulatory agencies such as the U.S. Food and Drug Administration (FDA). Thus, the implementation of standards is incredibly widespread within pharma and therefore knowledge cannot only be drawn from the wide pool of administrative resources but also from colleagues in the field.

However, in nutrition, though efforts have been made by regulatory agencies such as the European Food Safety Authority (EFSA) to create standards for food [4], and even CDISC has recently come out with a nutrition TAUG [5], the implementation of these standards to nutritional data is yet to be as widespread as in pharma. This could be attributed to the lack in requirements made by regulatory agencies but is also due to the fact that the data that is collected can be innately so different that standardisation is one tough cookie.

This paper will focus on our personal experiences across the aisles of pharma and nutrition and discuss certain examples of the differences between the two. At the end, viewpoints will be summarised, and lessons learned presented that will hopefully provide some food for thought.

## REGULATORY REQUIREMENTS

As mentioned in the introduction, there are regulatory agencies that govern the conduct of in-human experimentation and set the rules that are in place. They provide a guiding presence when it comes to the implementation of standards. In this chapter, we will explore the influence regulatory agencies can exert on the implementation of standards in industry and what regulation requirements are expected per industry.

**PHARMA**
There are many regulatory parties involved in a new drug application (NDA) process. The very nature of pharmaceuticals and its history has led to the introduction and implementation of many laws governing clinical conduct in research done with human subjects. The role of these agencies are essential and they are heavily involved in the study registration and submission proces. Therefore there is a massive incentive for pharmaceutical companies to comply and be up to date with standards set by these agencies if they wish to see their products reach the market. We see this with the ongoing interest within the pharmaceutcial industry to learn and implement standards as best as possible at conferences like PHUSE.

**NUTRITION**
Although there are agencies involved in nutritional studies, such as the FDA and the EFSA, there are often less (strict) requirements for nutritional studies. Within nutritional studies there is a high variability in the type of product studied. Nutrition studies can for example focus on medical food, infant nutrition, regular food products or dietary supplements. Besides the different type of nutrition studies, every country has its own regulations and a global consensus is lacking.

As described by Dwyer et al. countries may even differ in what they define as food.[6] For example, a dietary supplement is considered a food in the USA but may be considered a therapeutic good or even a controlled substance in another country. To make matters even more difficult there are also countries such as China and India which have their own regulations for traditional (herbal) medicine.

The extent to which regulatory agencies are involved in nutritional studies highly depends on the type of product studied. For example when studying medical nutrition or infant nutrition, the target population is more vulnerable and the product may potentially cause more harm than regular food products. Typically, these types of products are subject to more strict laws than regular food products such as fruits and vegetables.

Besides specialised products or specific components, a lot of nutritional trials tend to study regular food products such as fruits and vegetables or dietary patterns, such as the Mediterranean diet. As these food products are products humans have been eating for centuries, there is no need for the interference of regulatory agencies to ensure safety or efficacy. Furthermore, a lot of nutritional studies are for internal use and are conducted by hospitals and/or universities that have no “monetary” incentive for conducting these studies.

## CRF DESIGN
The case report form (CRF) is an integral and defining part of any clinical study. It translates the design and primary objectives of a study to the clinical set up and helps guide clinicians in the process of data collection. This protocol driven documentation also serves as the guiding force in the standardisation of the collected data if designed properly. In this chapter, we provide some examples per industry and illustrate how the CRF design plays a significant role in the easement of data collection and standardisation.

**PHARMA**
Oftentimes in pharmaceutical studies the CRFs used are electronic (so-called eCRFs), and they often have intrinsic data quality protection built in. They are designed to capture data in a standardised way, for example by means of checkboxes (see figure 1), pre-defined formats/entry fields and CRF sections that only become available when certain options were selected. Furthermore, CRFs are often characterised by a lack of free (format) text answers. Moreover, data collection that is crucial for the study is often pre-printed or pre-defined on the CRF. A prime example of this practice is in the pre-printing of units of measure.

Additionally, as studies in the pharmaceutical industry are bound to CDISC standards (except for legacy studies), the CRFs are often annotated based on the CDISC Study Data Tabulation Model (SDTM) standard, which is a requirement when submitting to the FDA. With this in mind CRF design is more focussed on capturing data in a way that is suitable for SDTM purposes.

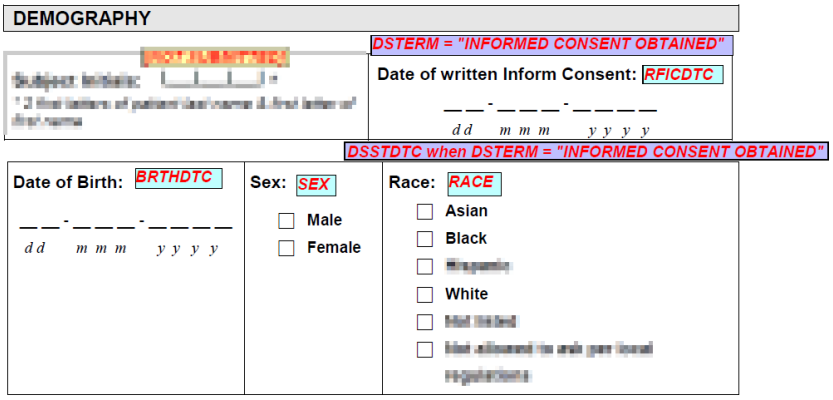


*Figure 1 – A depiction of a demography form of a CRF. Note the use of checkboxes and the pre-specified formats for the date entry fields.*

As one can see in figure 2, questions regarding the intake of certain substances (including supplements) can be asked and are sometimes incorporated. However these sections tend to be quite small and do not inquire after a patient's lifestyle unless that was part of an objective of the study.

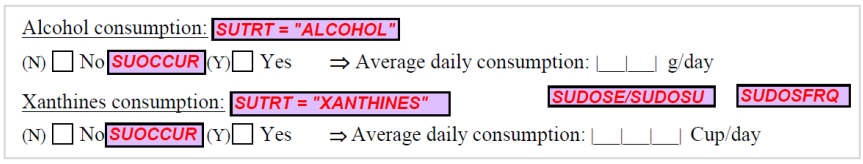


*Figure 1 - Section of CRF regarding substance use.*

Another example shown here (figure 3), is the collection of adverse event data with an eCRF (often used for primary and/or secondary safety outcomes) which also makes use of standardised entry fields.

| | |
|---|---|
| **Were any adverse events experienced?** AEYN Not submitted | ☐ Yes ☐ No <From NY codelist> |
| **AE Number:** | Choose an item. |
| **What is the adverse event term?** AETERM | Choose an item. |
| **Start Date:** AESTDAT | DD-MON-YYYY |
| **Is the adverse event ongoing?** AEONGO AEENTPT AEENRF | ☐ Yes <From NY codelist> |
| **End Date:** AEENDAT AEENDTC | DD-MON-YYYY |
| **Severity:** AESEV | ☐ MILD ☐ MODERATE ☐ SEVERE |
| **Was the adverse event serious?** AESER | ☐ Yes ☐ No <From NY codelist> |

*Figure 2 - Depiction of a section of the adverse event form of an eCRF with pre-defined entry fields.*

**NUTRITION**

CRFs in nutritional studies are often not designed with standardisation purposes in mind. In nutritional trials, a lot of data is collected in diaries and is stored as free text. Collection of study data in diaries may lead to difficulties in the mapping of timing variables such as --TPT (planned time point) if this is not taken into account while designing the CRF. The usage of free text is sometimes inevitable, however using free text fields may result in illogical and incorrect values for variables that could be easily standardised, such as country or sex, as shown in figure 4.

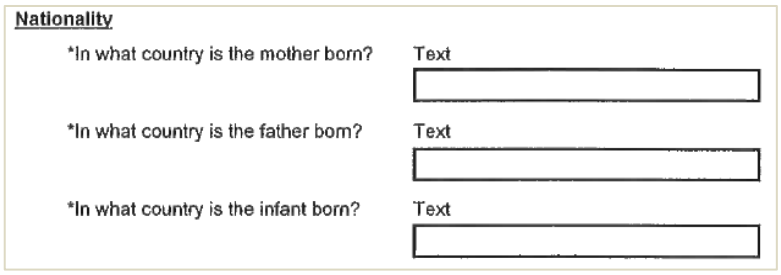


*Figure 3 - Example of a nutrition CRF section regarding the collection of country-related information.*

Data collection through questionnaires is not uncommon in nutritional studies. Figure 5 shows an example of such a questionnaire. What occurs often in nutritional studies are question flows, due to conditional questions in the questionnaire. These question flows might lead to difficulties when mapping these data. Is data of non-answered questions considered missing or is the answer to the questions automatically “no”, since the leading question was answered with “0 times in a usual day”?

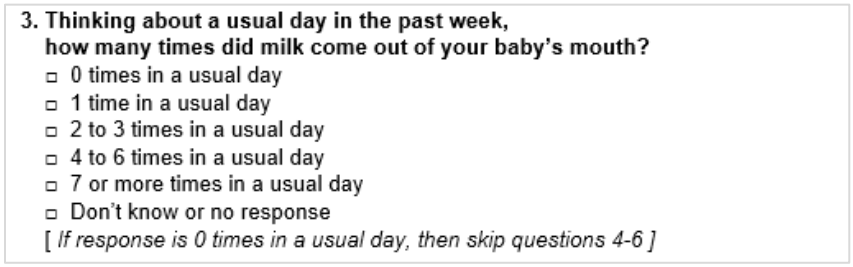


*Figure 4 - Example of a nutrition CRF section with conditional questions.*

To be able to clearly compare examples between the pharmaceutical industry and the nutritional industry, figure 6 is provided. In this figure the use of medication or nutritional supplements is captured. As one could see in the pharma example (figure 2), the use of other medication or supplements is also annotated and standardised on the CRF. However, in the nutrition study this is a free-text field. There is no specification of dosage and, in fact, there are multiple types of answers that can be given. Brand, type, but also duration and reason for taking supplements are examples of answers that are often given in these free-text fields.

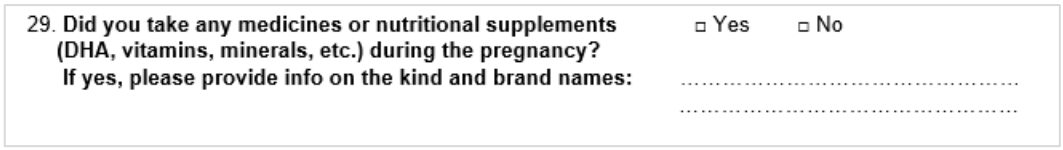


*Figure 5 - Example of a nutrition CRF section capturing usage of medication or nutritional supplements.*

The usage of diaries, free-text fields and conditional questionnaires is a big contrast to CRFs in the pharmaceutical industry. Free-text fields may lead to a lot of time and effort spent cleaning the data after data collection. This could be easily overcome by giving limited options on the CRF and during data entry. Although diaries and conditional questionnaires may lead to difficulties when mapping these data, sometimes these methods are inevitable. However, keeping these difficulties in mind while designing the CRF may ease the standardisation of these data afterwards.

## DATA COLLECTION AND MANAGEMENT

The previous chapter touches upon the significance of a well-designed CRF. The manner in which data is collected and managed will often dictate the likelyhood of post-data collection corrections or problematic data issues. Data collection and management is therefore, together with a proper CRF design, essential in successful standardisation. This chapter will continue to focus on observed differences between the two industries in this regard.

### PHARMA

Data collection and management in itself is a very well-established and well-catered part within a clinical trial. Data collection and management is either done by an internal business unit of a pharmaceutical company or is outsourced to a company (CRO) that is specialised in this area.

Usually the (e)CRF sections are filled in during a clinical visit by a ‘blinded’ clinician, as well as any assessments made. They are instructed in how to fill in the CRF and have completion guidelines to abide by which results in less

variance in data capture errors across different sites. Additionally, clinical data is captured with electronic data capture (EDC) software which often comes with pre-built-in data quality assessments. Data from these databases can be processed using validated software (e.g. SAS) or open source software (e.g R). Issues at clinical sites are overseen by data managers and resolved as quickly as possible resulting in less data issues when standardising.

In short, data collection and management is a well-regulated job within clinical trials and is handled by highly trained professionals with a background in data and clinical research resulting in an efficient handling of data.

**NUTRITION**

Data collection in nutritional studies often does not only include data about food intake itself. Factors such as sunlight, exercise, sleep habits or living in an urban or rural environment may highly influence the outcome studied, this data is captured to be later used in analysis as confounders or effect modifiers.
One major nutrition challenge is the different units in which food intake can be measured. One could choose to study vitamin intake, food group intake (e.g. fruits) or even compare complete diets. Furthermore, there are different methods of capturing food intake, for example a food frequency questionnaire, 24-hour recalls or food diary each having their advantages and disadvantages one over the other.

Data collection can differ greatly amongst nutritional studies, depending on the researchers and the organisation they are working for. Data is sometimes captured with EDCs, but can also be captured in Excel or SPSS, which are commonly used by hospitals and universities. Data analysis is often performed using SPSS for its user-friendly point-and-click approach, or R for being open-source software. Because there is less interference of regulatory agencies, software does not need to be validated or approved of. Also, as budget tends to be lower in nutrition studies than in pharmaceutical trials, the preferred choice is for easy accesible, user-friendly and less expensive software programs than in the pharmaceutical industry.

In contrast to the pharmaceutical industry, where we see that different stages of the trial are executed by different parties, this is mostly not the case in nutrition studies. Data collection, data management and data cleaning is often performed by principal investigators and research assistants in nutritional trials. Contrary to the pharmaceutical industry, all these processes are often done by the same person.

## DATA STANDARDISATION

Though there is a massive amount of resources to turn to for guidance when it comes to standardisation and common or accepted practices for pharmaceutical data, there remains a gap when it comes to standardising nutritional data. In this chapter, we shall explore some of the underlying reasons behind this phenomena and expand on the differences in the approaches taken in each industry when it comes to data standardisation.

**PHARMA**

Since its foundation in 1997, CDISC has always strived to establish data standards for all types of information collected during clinical trials to the benefit of the medical and pharmaceutical world. Their mission statement and purpose has served the pharmaceutical industry well and its ‘pharmaceutical’ stake is very well represented, especially in the CDISC SDTM standard. The nature of variables used within this model cater extensively to pharmaceutical type data.

Not just the many SDTM domains reflect this ‘pharmaceutical’ identity, it is also represented in the amount of TAUGs available for clinical trials of (bio)pharmaceuticals. This extensive pool of examples and guiding documents greatly aides the standardisation ventures of those handling pharmaceutical data.

**NUTRITION**

When it comes to standardisation a lot of SDTM domains are mainly focussed on the pharmaceutical industry. As the SDTM standards are based on pharmaceutical data, it is found that some variables are typically used for pharmaceutical data, but are not that commonly used for nutritional data. This creates some obstacles to overcome when trying to fit nutrition data into CDISC standards. One way of how this is reflected is in which variables are considered as being required, expected and permissible. Another way is the controlled terminology that contains codelists that are sometimes non-extensible, however do not always cover the terminology needed for nutritional studies.

Also, in contrast to a lot of findings domains in pharma, in nutritional studies a lot of “findings about” domains are used as the regular findings domains are not always fit for nutritional data.

**NUTRITION TAUG**

With the release of the TAUG for nutrition in September 2019, guidance was created on how to standardise nutrition data. This TAUG provides insight on how to map data related to infant nutrition, stool samples and stool assessments. There are some topics however, that are not covered yet. An example is the differences in mapping food, or product data and macro- or micronutrient data and their hierarchy. Guidelines on how to handle this "parent-child" data is still under construction.

The same applies for the modelling of --ORRES/--STRESC when the CRF is not designed with SDTM standards and controlled terminology in mind. It has to be noted that the CRF examples in the nutrition TAUG are based on annotated CRFs which collect the data nearly in SDTM format already. As we have discussed earlier in this paper, this would be best practice for nutrition studies, however this is not always the case.

**DIFFERENCE IN DOMAINS - THE EXPOSURE DOMAIN (EX) AND THE MEAL DOMAIN (ML)**

The exposure domain is a great example of how CDISC standards accommodate pharmaceutical data. For example, note the use of "DOSE" and "DOSFRM", which reflects the expectation of pharmaceutical data (see table 1).

For the nutritional data we see the yoghurt with the probiotics (op. cit. abstract) in the EX domain (see table 2), comparable to the exposure in the pharma study. We see the other intake being displayed in the ML (Meal) domain (see table 3). Exposure is often collected rather detailed, this is also the case in nutritional studies, making variables as --DOSE and --DOSEU easier to fill out. However, if we look at the other meal data, we see for instance two slices of bread with jam. This type of data is hard to collect as detailed as what is being done for pharma. Moreover, the use of variables such as --DOSFRM and --ROUTE may be useful in some nutritional studies, but seem redundant for most.

Another aspect that can make nutritional data harder to map with regard to the EX/ML domain can be found when we look at an example of infant studies where mixed feeding is allowed. In the example of this paper, it is very clear that yoghurt with added probiotics, although being a food product, is the exposure. Hence, the yoghurt will be mapped to the EX domain, while all the other intake (the sandwiches with jam,grapefruit juice and water) are mapped to the ML domain. However, in infant studies, some study designs allow for the infant to receive mixed feeding, often meaning that the infant receives both infant formula and breastfeeding. In this case it is good to think about clear rules when something is considered being exposure data or meal data and to map this consistently throughout the study or studies.

*Table 1 - Example of an EX domain for the pharma study.*

| STUDYID | DOMAIN | USUBJID | EXTRT | EXDOSE | EXDOSU | EXDOSFRM | EXROUTE |
|---|---|---|---|---|---|---|---|
| CDISC | EX | CDISC-01 | DRUGA100 | 5 | mg | SUSPENSION | ORAL |

*Table 2 - Example of an EX domain for the nutritional study.*

| STUDYID | DOMAIN | USUBJID | EXTRT | EXDOSE | EXDOSU | EXDOSFRM | EXROUTE |
|---|---|---|---|---|---|---|---|
| CDISC | EX | CDISC-01 | YOGHURT WITH ADDED PROBIOTICS | 150 | ml | SUSPENSION | ORAL |

*Table 3 - Example of an ML domain for the nutritional study.*

| STUDYID | DOMAIN | USUBJID | MLTRT | MLDOSE | MLDOSU | MLDOSFRM | MLROUTE |
|---|---|---|---|---|---|---|---|
| CDISC | ML | CDISC-01 | SANDWICH | 2 | SLICE | | ORAL |
| CDISC | ML | CDISC-01 | JAM | | | | ORAL |
| CDISC | ML | CDISC-01 | GRAPEFRUIT JUICE | 1 | GLASS | | ORAL |
| CDISC | ML | CDISC-01 | WATER | 1 | GLASS | | ORAL |

**DIFFERENCE IN DOMAINS - THE DRUG ACCOUNTABILITY DOMAIN (DA)**

The Drug Accountability dataset, though currently perfect for pharmaceutical data, could be updated to fit more than only data regarding the accountability of a drug (see table 4 and 5). Drug Accountability is not a domain where one would map compliance of the nutritional intervention to. For this reason, the nutrition TAUG now refers to the DA domain with an updated description of "Product Accountability" in order to extend the use of the domain. This shift in SDTM domains would allow for a greater variability in product data that can be accommodated in the SDTM standard.

*Table 4 - Example of a DA (Drug Accountability) domain for the pharma study.*

| STUDYID | DOMAIN | USUBJID | DATESTCD | DATEST | DACAT | DAORRES | DAORRESU |
|---|---|---|---|---|---|---|---|
| CDISC | DA | CDISC-01 | DISPAMT | Dispensed Amount | Study Product | 30 | CAPSULES |
| CDISC | DA | CDISC-01 | RETAMT | Returned Amount | Study Product | 10 | CAPSULES |

*Table 5 - Example of a DA (Product Accountability) domain for the nutrition study.*

| STUDYID | DOMAIN | USUBJID | DATESTCD | DATEST | DACAT | DAORRES | DAORRESU |
|---|---|---|---|---|---|---|---|
| CDISC | DA | CDISC-01 | DISPAMT | Dispensed Amount | Study Product | 300 | mL |
| CDISC | DA | CDISC-01 | RETAMT | Returned Amount | Study Product | 150 | mL |
| CDISC | DA | CDISC-01 | DISPAMT | Dispensed Amount | Study Product | 2 | CUP |
| CDISC | DA | CDISC-01 | RETAMT | Returned Amount | Study Product | 1 | CUP |

## SUMMARY

It has been shown that major differences between these two industries arise even before the study actually starts. The manner in which CRF's are designed in pharma are a big influence on the standardisation of the data collected. Collecting data in more structured fields would ensure higher data quality and by letting data management focus more on standardisation, the standardisation process can be made easier for nutrition studies.

## DISCUSSION

Though pharmaceutical studies may inquire after certain food or substance(s) of interest, there is no further incentive to document other "lifestyle" factors that could influence the participants behaviour and subsequently perhaps their outcomes, unless that was part of a study objective (e.g. diet related inquiries in a diabetic study population). When it comes to standardisation there is more to learn from pharma by nutrition than the other way around.
This is mostly due to the fact that the sense of urgency is lower in nutrition studies, because standardisation is not mandatory as per EFSA requirements. Although there is a shift in mentality with regards to nutrition studies, fully adopting standards will take time (and thus money). To help aid standardisation, it is important to open the discussion on the appropriateness of the SDTM "fit" for nutritional data.

The hope is, with the release of the nutritional TAUG and presentations on how to implement these standards at conferences like PHUSE and CDISC, the nutrition industry will be stimulated and supported to standardise their data more and more.

## CONCLUSION

In conclusion, pharma is more focussed on standardising data. However, there is a notable shift happening in nutritional studies. Standardisation of nutrition studies is not only feasible but has certain advantages, most notably;
1) to be able to expand on the sustainability of the data (i.e. can be used in pooling efforts)
2) the possibility to be able to reproduce the data which is exceedingly beneficial to quality control (QC) and auditing.

This paper provided a short overview on personal viewpoints regarding the differences and similarities between the industries of pharma and nutrition and will help advance discussion regarding this subject that will aid both industries.

## REFERENCES

[1] A. Bhatt, “Evolution of clinical research: a history before and beyond james lind.,” *Perspect. Clin. Res.*, vol. 1, no. 1, pp. 6–10, Jan. 2010, Accessed: Aug. 06, 2020. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/21829774.

[2] D. Zeilstra, J. A. Younes, R. J. Brummer, and M. Kleerebezem, “Perspective: Fundamental limitations of the randomized controlled trial method in nutritional research: The example of probiotics,” *Adv. Nutr.*, vol. 9, no. 5, pp. 561–571, Sep. 2018, doi: 10.1093/ADVANCES/NMY046.

[3] D. Mozaffarian, I. Rosenberg, and R. Uauy, “History of modern nutrition science-implications for current research, dietary guidelines, and food policy,” *BMJ*, vol. 361, Jun. 2018, doi: 10.1136/bmj.k2392.

[4] “The food classification and description system FoodEx 2 (revision 2),” *EFSA Support. Publ.*, vol. 12, no. 5, Mar. 2017, doi: 10.2903/sp.efsa.2015.en-804.

[5] “Nutrition | CDISC.” https://www.cdisc.org/standards/therapeutic-areas/nutrition (accessed Aug. 06, 2020).

[6] J. T. Dwyer, P. M. Coates, and M. J. Smith, “Dietary supplements: Regulatory challenges and research resources,” *Nutrients*, vol. 10, no. 1. MDPI AG, Jan. 04, 2018, doi: 10.3390/nu10010041.

## ACKNOWLEDGMENTS

We would like to acknowledge our wonderful colleagues who helped us finalise this paper by providing constructive feedback and encouragement along the way. In particular, we would like to mention Lieke Gijsbers and Sofia Vale for proof-reading and providing critical talking points.

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Louella Schoemacher & Fatima Kassim
OCS Life Sciences – OCS Consulting B.V.
Ruwekampweg 2G
‘s-Hertogenbosch / 5222 AT
**+31 (0)73 523 6000**
sasquestions@ocs-consulting.com
https://ocs-consulting.nl/ocs-life-sciences/