The consumption of varying frequencies, varieties, and quantities of fruits & vegetables and pulses, nuts & seeds among children 6-23 months of age and their association with dietary and health outcomes: a systematic review and meta-analysis

Leila Harrison[1], Christina Oh[1], Kimberly D. Charbonneau[1], Aatekah Owais[1], Emily C .Keats[1] and Zulfiqar A. Bhutta[1]

[1]Centre for Global Child Health, The Hospital for Sick Children, Toronto, Canada

Table of Contents

ABSTRACT

Background

Adequate nutrition during infancy is essential to the development of each child's full potential later in life. The window between birth and two years of age is well known to be a critical period for the physical and cognitive growth of infants. As of 2020, among children under the age of five, undernutrition has been associated with up to 45% of deaths and 38.3 million children are overweight or obese. Suboptimal breastfeeding and complementary feeding practices, when combined with infectious diseases, are the leading causes of malnutrition during these first two years of life and continue to be highly prevalent, especially in low-and middle-income countries (LMICs).

The World Health Organization (WHO) recommends feeding a variety of complementary foods, including fruits and vegetables (FV), animal-source foods (ASF), and nuts, pulses, and seeds (NPS) in order to meet nutritional demands during this period of rapid growth and development. The daily consumption of FV and NPS in particular have been recommended due to their association with health benefits. More specifically, FV have been found to be an important food group in reducing vitamin A deficiencies, as well as many other vitamin needs. NPS offer a significant source of protein and iron (when combined with vitamin C-rich foods) and are particularly important in LMICs where ASF that tend to be costly are not as accessible.

This review aims to assess the health, development, and dietary effects of more frequent or more varied consumption of FV and/or more frequent or greater amounts of NPS, compared to less frequent, varied, or lower amounts of consumption, during the complementary feeding period.

Objectives

Primary objectives: To compare i) the effect of more frequent versus less frequent consumption of fruits, ii) the effect of more frequent versus less frequent consumption of vegetables, iii) the effect of more frequent versus less frequent consumption of NPS, iv) the effect of more varied versus less varied consumption of fruits, and v) the effect of more varied versus less varied consumption of vegetables, and vi) the effect of more amounts versus less amounts of NPS consumption among infants 6-23 months on dietary and health outcomes later in life.

Secondary objective: To determine differences between processed/commercial versus fresh/home-prepared fruits and vegetables on dietary and health outcomes.

Search Methods

The search strategy was formulated using the PICO methodology based on medical subject headings and key words and run in the following electronic databases: MEDLINE, Embase,

CINAHL, African Index Medicus, LILACS, Cochrane CENTRAL, and eLENA (WHO). Searches were also conducted for non-indexed, grey literature using Google Scholar and select organizational websites, as well as reference lists of all relevant systematic reviews captured in the search strategy. All search results were screened in duplicate.

Selection Criteria

The criteria for considering studies into this review were as follows: healthy male and female children aged 6-23 months from any LMIC or high-income country; consumption of FV and/or NPS that has been reported in terms of frequency, variety, or amount; and the study must have measured a primary or secondary outcome of interest.

Exclusion criteria: recruitment of “unhealthy” child populations; and qualitative studies or reviews.

Data collection and analysis

Four review authors extracted data into a piloted excel sheet and completed risk of bias assessments for each included study (in duplicate). Any disagreements were resolved through discussion or by a third reviewer. Individual studies were critically appraised using the Cochrane Risk of Bias-2 (ROB-2) tool for randomized controlled trials and cluster-randomized controlled trials, the NIH tool for observational cohort and cross-sectional studies, and the ROBINS-I Tool for non-randomized studies.

All experimental and observational study data were analyzed separately. Conducting meta-analysis was not possible for any outcome, given the heterogeneity across studies in food group item consumed, outcome metrics, and frequency categories. As such, a narrative synthesis was conducted across all exposures and outcomes.

Results

Study characteristics: We included six studies that analyzed data from 23,346 children that ranged from 6 months to 23 months of age at time of enrollment. The included studies took place in Senegal, Brazil, Indonesia, China, and Norway (n=2). All of the included studies were observational by design, where five of the included studies were cross-sectional studies and one was a longitudinal cohort study, and thus we were unable to establish causal inference with outcomes. All six studies looked at frequency of consumption of vegetables, five studies looked at frequency of consumption of fruits, and two studies looked at frequency of consumption of NPS. It should be noted that none of our included studies looked at the association between differing varieties or amounts of FV or NPS on dietary and health outcomes. Furthermore, no studies have looked at NPS consumption beyond legumes. The following outcomes were reported: wasting, underweight, stunting, linear growth, height-for-age z-score (HAZ), iron status, anemia, and changes and stability in FV intake later in life.

Certainty of evidence: Overall, we rated all six studies as having a very low certainty of evidence.

Key results: We found one very low certainty cross-sectional study conducted in 3 sub-districts of Aceh Besar District, Indonesia that reported no association between varying frequencies of consumption of dark green leafy and orange vegetables, fruits, or legumes and wasting, underweight, and stunting. We found one very low certainty cross-sectional study conducted in a mixture of urban and rural locations in China, that demonstrated an association between less frequent consumption of FV and greater prevalence of stunting and underweight. We found one very low certainty cross-sectional study conducted in rural Senegal that points to an association between lower HAZ and less frequent consumption of vegetables/leaves and fruit. The same study reported that more frequent fruit consumption was positively associated with linear growth. However, contrary to our expectation, this study also showed that frequent consumption of vegetables had an inverse relationship with linear growth. We found one very low certainty cross-sectional study conducted in urban Norway that reported an association between greater consumption of vegetables and greater likelihood of having low ferritin values (<20μg/L). However, these findings were inconsistent, as this relationship was not seen for those eating vegetables at an even higher frequency, and there was no significant relationship observed between the feeding frequency of vegetables and the likelihood of having very low iron stores (ferritin <15μg/L). We found one very low certainty cross-sectional study conducted in urban Brazil that reported no association between consumption of dark green vegetables, fruit (consumption in past 24 hours vs not) or beans and anemia prevalence. However, when a different frequency metric was used for fruits, specifically daily vs less than daily, less fruit consumption was associated with higher anemia prevalence. Lastly, one very low certainty longitudinal study conducted in a mixture of urban and rural locations in Norway demonstrated that overall FV consumption at 18 months was positively associated with overall FV consumption at both 36 months and 7 years of age.

Author's conclusions

The findings from this systematic review aimed to inform the development of the updated WHO infant and young child feeding guidelines along with country-specific policies and programmes relating to complementary feeding. However, this review found a lack of relevant and good quality data, where no causal relationship could be established between the exposures and outcomes reported, on which to base practice and policy at this time. Instead, it has highlighted a major gap in the evidence, underscoring the need for more research of higher quality to be conducted related to this topic.

BACKGROUND

The Problem, Condition, or Issue

Adequate nutrition during infancy is essential to the development of each child's full potential later in life.[1] The period between birth and two years of age is known as a "critical window" as it is one of the most important periods of physical and cognitive growth for infants and a time where their development is exceptionally vulnerable to insults such as those resulting from poor diet and disease.[1,2] Longitudinal studies have consistently shown that poor diet during these formative years can lead to morbidity, mortality, and delayed motor and mental development in the short-term.[1,3] In the long-term, impaired intellectual performance, work capacity, poor reproductive and overall health outcomes can lead to significant loss of human capital in the later years of life.[1,4,5]

Malnutrition in children is comprised of both undernutrition (stunting and wasting) and overnutrition (overweight and obesity). Both under- and overnutrition may be associated with micronutrient deficiencies. Undernutrition is associated with up to 45% of deaths among children under the age of 5; as of 2020, 47 million children under the age of 5 are wasted, 14.3 million are severely wasted, and 144 million are stunted.[6] Additionally, 38.3 million children under the age of 5 are overweight or obese.[6] Of the major micronutrient deficiencies, vitamin A, zinc, iron, and iodine are of most concern, as they have been linked to the largest proportions of years of life lost and disability adjusted life years.[7,8]

Poor breastfeeding and complementary feeding practices, in combination with infectious diseases, are the principal causes of malnutrition during the first two years of life.[1] Exclusive breastfeeding is recommended until an infant is 6 months of age. Complementary feeding refers to the period in which breastmilk only is no longer considered sufficient to meet the nutritional requirements of the growing infant and thus other foods and liquids are needed alongside breastmilk to achieve optimal growth and development.[1] There are two types of complementary foods: specially prepared foods and typical family foods that have been modified for the infant.[9] For example, a caregiver may specially prepare a porridge dish for the infant, while the rest of the family consumes a nut stew. When the infant is a bit older, the caregiver may mash the nut stew into a different consistency and texture, making it easier for the infant to consume the family food.[9] Although breastfeeding may continue past 2 years of age, complementary feeding is typically targeted to infants from 6 to 23 months, after which they are expected to consume family foods. While the numerous benefits of continued breastfeeding (>6 months) are well understood, suboptimal complementary feeding practices are still highly prevalent, especially in low- and middle-income countries (LMICs).[10-14]

The Intervention

Based on an average intake of breastmilk among healthy infants, the energy needs from complementary foods are estimated to be about 200 kcal per day for infants aged 6-8 months,

300 kcal per day for infants 9-11 months, and 550 kcal per day for infants 12-23 months.[1] This translates into a meal frequency of 2-3 times per day for the youngest infants (6-8 months) and 3-4 times per day through the rest of the complementary feeding period.[1] For non-breastfed infants, the energy needs from complementary foods are estimated to be about 600 kcal per day for infants 6-8 months, 700 kcal per day for infants 9-11 months, and 900 kcal per day for infants 12-23 months.[15] For non-breastfed infants, the meal frequency is defined as 4-5 times per day throughout the entire complementary feeding period (6-23 months)[15,16] However, needs will vary depending on breastmilk intake, growth rates, and other factors, such as illness and environmental conditions. The World Health Organization (WHO) also recommends feeding a variety of complementary foods, including fruits and vegetables (FV), animal source foods (e.g., meat, poultry, dairy products, eggs), as well as nuts, pulses, and seeds (NPS) in order to meet nutritional demands during this period of rapid growth and development.[1,17]

Complementary foods need to provide enough energy, protein, and micronutrients to ensure that in combination with breast milk, the infant is meeting all their energy and nutrient needs for optimal growth and health. Given the low quantity of foods that are typically consumed, it is especially important that they are nutrient-dense. While vegetarian diets are not advised because of their inability to provide adequate amounts of some micronutrients, the daily consumption of FV and NPS is recommended due to its association with health benefits.[1,18] Notably, vitamin A-rich fruits and vegetables are important for reducing vitamin A deficiencies in infants and have been found to meet many other vitamin needs.[1,19] Similarly, NPS can offer a significant source of iron when consumed with vitamin C-rich foods.[18]

Typically, an infant will consume a staple food (e.g., cereals, roots, or starchy fruits) that provides them with most of their energy needs. In addition to this staple, a mixture of foods such as NPS, foods from animals, dark-green leaves or orange-coloured FV, and oils, fats, and sugars need to be consumed in order to fill the remaining energy and nutrient requirements. FV in particular play a major role in filling this nutrient gap. For example, an infant 6-23 months old only needs to consume 1 tablespoon of sweet potato in order to meet their vitamin A needs for one day.[9]

It is important that consumption of any complementary food, including FV and NPS, is not excessive to the point that breastfeeding frequency is reduced. For example, a social marketing campaign in Guatemala found that promoting complementary feeding 5 times per day had the unintended consequence of displacing necessary breastfeeding, whereby average daytime breastfeeding reduced from 6.9 feedings to 3.7 feedings from pre-intervention to post-intervention time points ($p=0.01$).[20] In addition, food preparation and feeding a young infant 5 times per day takes a considerable amount of effort and time from the caregiver, leading them to potentially use the same prepared food from one meal to the next. This may increase the risk of microbial contamination, as well as reduce dietary diversity, and needs to be considered when discussing appropriate meal frequency during complementary feeding.[1]

This review assesses whether more frequent or more varied consumption of FV and/or more frequent or greater amounts of NPS, compared to less frequent, varied, or amounts of

consumption, is beneficial for the growth and development of an infant during the complementary feeding period.

How the intervention might work

FV are an essential source of nutrients and vitamins that are vital to health and growth, such as potassium, folate, fiber, vitamin A, vitamin C, vitamin K, and many phytochemicals.[21-22] For this reason, caregivers are recommended to offer FV when complementary foods are introduced to a milk-based diet.[23] Repeated exposure to FV during infancy has been associated with better acceptance and improved intake later in childhood.[17,24-28] A prospective longitudinal study from England with 7,821 participants showed that exposure to home-prepared FV during the early stages of complementary feeding was positively associated with increased frequency and variety of intake of these foods later in life at age 7.[29] Similarly, data from another longitudinal study carried out in the US with 1,542 participants found that infrequent consumption of FV during the first year of life was associated with low FV intake at 6 years of age.[23]

Despite our understanding of the benefits of consuming FV during this critical window, several studies have shown that at the population-level, FV intake during the complementary feeding period remains low.[23,30-31] Instead, FV are replaced with high energy, nutrient-poor, and processed items such as sweet foods and sugary drinks.[23,30-31] This is not only associated with negative consequences on short and long-term health and nutrition outcomes, as low FV intake is a major risk factor for micronutrient deficiencies, mortality, and chronic diseases (e.g., cardiovascular disease, cancers, and type-2 diabetes),[32] but may also reduce the acceptance of these foods and negatively influence eating behaviors later in life.[24-28]

NPS provide important macronutrients (energy, protein, essential fats, and fiber) and micronutrients (e.g., iron, zinc, etc.). Iron is the largest nutrient gap in infants 6-23 months old, where 43% (273 million) of children globally less than 5 years old were found to be anemic in 2011.[33] Iron, especially bioavailable iron, is most abundant in animal sourced foods such as meat, poultry, or fish. However, NPS are another important source of iron when combined with other foods rich in vitamin C that allow for improved absorption of iron by the body.[18] In addition to an infant's iron needs, NPS are also a good source of protein, an important macronutrient. NPS are a particularly important food group in low-income populations, as they can add nutritional value to diets when animal source foods, which tend to be costly, are not accessible. Their long storage life and relatively low cost make them a key source of protein, energy, and micronutrients, particularly in low and middle-income countries.

Concerns around allergies can be a reason why commonly allergenic foods, such as nuts (e.g., peanuts), are not introduced early to a child. However, there is accumulating evidence to suggest that delays in introduction may actually promote the development of, versus prevent, the allergy.[17,34] Another common concern around the provision of NPS to infants includes choking; though, parents are advised to puree or mash the nut or seed into a paste, thereby allowing the child to consume the food safely while receiving the nutritional value that NPS provides.

Why it is important to do this review

Complementary feeding is a critical time of transition for an infant, and inappropriate complementary feeding practices are associated with adverse short- and long-term health and nutrition outcomes. Studies from across the globe consistently show that most infants do not consume the recommended amounts of FV and NPS, contributing to widespread nutrient and vitamin deficiencies.[10-14, 35-37] However, to date, there have been no reviews assessing the effect of more frequent or varied consumption of FV and/or more frequent and greater amounts of NPS, compared to less frequent, varied, or amounts of consumption, during the complementary feeding period on dietary and health outcomes. What literature does exist reports on incidence of FV and/or NPS intake, or interventions to improve consumption of these foods during the complementary feeding period (e.g., educational interventions aimed at improving complementary feeding practices). Specific guidance around the optimal frequency and variation in types of FV and NPS that should be consumed is lacking. We therefore seek to assess the dietary and health outcomes that are associated with more versus less frequent, varied, and amounts of consumption of these food groups. Findings from this review will help to inform the development of the updated WHO Infant and Young Child Feeding (IYCF) guidelines, along with future policies and programmes that relate to complementary feeding. This will have important implications for the health and wellbeing of infants, with lasting effects throughout the life course.

OBJECTIVES

The primary objectives of this systematic review are to compare: i) the effect of more frequent versus less frequent consumption of fruits, ii) the effect of more frequent versus less frequent consumption of vegetables, iii) the effect of more frequent versus less frequent consumption of NPS, iv) the effect of more varied versus less varied consumption of fruits, and v) the effect of more varied versus less varied consumption of vegetables, and vi) the effect of more amounts versus less amounts of NPS consumption among infants 6-23 months on dietary and health outcomes later in life.

Secondary objectives of this review are to determine differences between processed/commercial versus fresh/home-prepared fruits and vegetables on dietary and health outcomes.

METHODS

PICO Criteria for considering studies for this review

Types of Participants

Healthy male and female children aged 6-23 months, living in any low-, middle-, or high-income country. Studies were excluded if the infant population was considered unhealthy and infants were recruited based on this criterion. This includes, but is not limited to, infants with acute or chronic conditions/diseases, such as malnutrition, diarrhea or human immunodeficiency virus (HIV), and infants born preterm, small-for-gestational age (SGA) or low birthweight (LBW). Though our aim was to include healthy infants under two years of age, given the high burden of some of these conditions in certain populations globally, we recognize that malnourished infants and infants born preterm, SGA, or LBW are most likely included in the overall sample.

Types of Interventions & Comparators

Studies were eligible for inclusion if they: i) measured and reported data on consumption practices (reported as frequency, variety, and/or amount of food consumed) during the complementary feeding period; ii) examined consumption of FV or NPS (e.g., legumes, beans, lentils, peas); and iii) reported a relevant primary or secondary outcome (noted below).

We aimed to define frequency, variety and amount exposures for FV and NPS as follows:

- Frequency: none, once daily, 2-3 times per day, 3-4 times per day, >4 times per day, 1-3 times weekly, 4-6 times weekly, served either as meals or snacks.
- Variety: one type of food item per day, 2 different types per day, 3 different types per day or, >3 different types per day. These same variety exposures were evaluated on a per week basis as well.
- Amount, based on energy requirements by age:
    - 6-8 months: <137 grams (g) per day, 137-187 g/day, >187 g/day, OR <200kcal/day, ≥200 kcal/day
    - 9-11 months: <206 g/day, 206-281 g/day, >281 g/day, OR 9-11 months: <300 kcal/day, ≥300 kcal/day
    - 12-23 months: <378 g/day, 378-515 g/day, >515 g/day, OR <550 kcal/day, ≥550 kcal/day

The type of milk provided during the complementary feeding period was not a consideration for study selection (e.g., breast milk, animal milk, infant formula, or mixed).

The comparison for frequency, variety, and amount exposures for FV and NPS foods were as follows:

- Less frequent or varied consumption of fruits and/or vegetables
- Less frequent consumption or lower amounts of nuts, pulses, and/or seeds

Types of Outcome Measures

Primary outcomes:

1. Subsequent consumption of food items across the two food groups at 1 year, 2 years, 3 years, 4 years, 5 years, and beyond 5 years of age.
2. Nutrient adequacy (e.g., protein intake and quality, micronutrient intake, choline and essential fatty acids) or nutrient excess (e.g., saturated fat, protein, sodium, fibre/phytate by type, fats), as reported by the study authors.
3. Nutrient status at study endline (blood concentration of vitamin A, vitamin C, vitamin D, vitamin E, B vitamins, zinc, iron, folate, selenium, lutein, carotenoids, iodine, fatty acids (omega 3 and 6); anemia; antioxidants.
4. Anthropometric outcomes, including stunting (height-for-age z-score < -2 SD), wasting (weight-for-age < -2 SD), overweight (weight-for-height > 2 SD), and obesity (weight-for-height > 3 SD), as defined by the WHO Growth Standards.[38]
5. Anthropometric indices as continuous outcomes were also included and evaluated.
6. Child development (as defined by authors).
7. Contaminants within foods consumed (e.g., aflatoxins), as reported by the study authors.
8. Displacement of other foods/dietary adequacy, as reported by study authors.

Secondary outcomes:

1. Food/taste preferences later in life.
2. Markers of lipid profiles (e.g., total cholesterol, triglyceride, low-density lipoprotein, high-density lipoprotein) at endline or latest follow up.
3. Markers of inflammation (e.g., C-reactive protein, plasminogen activator inhibitor-1) at study endline or latest follow-up.
4. Markers of gut health (e.g., Bifidobacterium, Clostridia, short chain fatty acids, environmental enteric dysfunction, microbiome) at study endline or latest follow-up.
5. Adverse effects, as reported by the study authors.
6. Morbidity (infectious).
7. Food-borne illness (not related to storage and handling).
8. Bone health, as defined by authors.
9. Oral health, as defined by authors.
10. Adverse events (e.g., choking).

According to the GRADE framework, eight highest ranked outcomes were included and considered for quality of evidence and strength of recommendation.

Table 1. Relative Importance of Study Outcomes

Outcome	Score	Importance

Wasting	8	Critical
Underweight	8	Critical
Stunting	8	Critical
Anemia	8	Critical
HAZ	7	Critical
% Stability and change in vegetable/fruit consumption	7	Critical
Linear growth	5	Important
Ferritin <20µg/L	5	Important

Types of Settings

All settings within any low-, middle-, or high-income country were considered for this review.

Types of Studies

The types of studies included were primary studies of experimental design and observational design. Eligible study designs included randomized controlled trials (RCTs; cluster or individually randomized), non-randomized controlled trials, cohort studies (prospective and retrospective), cross-sectional studies, and case-control studies. Qualitative studies and reviews were excluded from this review.

Search Methods for Identification of Studies

Reporting and Protocol
The protocol for this review was registered with the International Prospective Register of Systematic Reviews (PROSPERO#: 203655).

Electronic searches

Our search strategy was formulated using the PICO methodology based on medical subject headings (MeSH) and key words (see appendix A for Medline search strategy). The search strategy was designed to capture studies evaluating the two food groups of interest (FV and NPS), along with animal-source foods (ASF). Studies were screened together until the analysis stage, at which point included studies were analyzed separately for each food group. This report will focus on the FV and NPS. Synthesis and findings from the ASF studies will be reported in a separate report.

Electronic searches were conducted in the following databases:

- MEDLINE
- Embase
- CINAHL
- African Index Medicus (AIM)
- LILACS

- The Cochrane Central Register of Controlled Trials (CENTRAL)
- eLENA (WHO)

Searches were also conducted in non-indexed, grey literature using Google Scholar and select organizational websites including UNICEF, Nutrition International, the Global Alliance for Improved Nutrition, Helen Keller International, and the CDC. These organizations were chosen based on their relevant work in the areas of infant and child diet and nutrition. There were no restrictions on publication date or language. The date of the final search was October 7, 2020.

Searching other sources

We searched the reference lists of all relevant systematic reviews captured during our electronic search for other studies that may not have been captured through this search strategy. In addition, the WHO's Guideline Development Group reviewed our list of included studies and provided suggestions for additional studies to screen.

Data Collection and Analysis

Selection of studies

Studies for title and abstract screening and full-text screening were managed using Covidence, a web-based software platform for systematic reviews. At both title/abstract and full-text screening stages, four review authors (CO, KC, KR, LH) independently scanned and screened all records retrieved by the searches for relevance based on selection criteria (see Table 2). Any disagreements were resolved through discussion or by a fifth review author when necessary. The selection process can be viewed in a flow diagram (see Figure 1).

Table 2. Inclusion and Exclusion Criteria

Inclusion Criteria
● Low-, middle-, or high-income country
● Healthy male and female children aged 6-23 months who have consumed fruits/vegetables, or nuts/pulses/seeds
● Types of interventions   ○ Consumption must have been reported in terms of frequency or variety for fruits/vegetables; and frequency/amount for nuts/pulses/seeds

● Relevant study designs:
  ○ Randomized controlled trials (RCTs)
  ○ Non-randomized controlled trials (NRTs)
  ○ Cohort studies
  ○ Case-control studies
  ○ Cross-sectional studies

● Study measures a primary or secondary outcome of interest

Exclusion Criteria
● Recruitment of “unhealthy” child populations, including but not limited to acute or chronic conditions/diseases, genetic diseases, malnutrition, diarrhea, or HIV
● Irrelevant study designs: reviews, qualitative studies

Data extraction and management

Four review authors (CO, KC, KR, LH) independently extracted data from each included study onto a standardized data extraction form in Excel that had been piloted. All studies were matched between the four review authors, and any disagreements or discrepancies were resolved through discussion, or by a fifth independent reviewer.

We extracted the following information from each included study: source (e.g., contact details); study characteristics (e.g., study design, location of study, years of data collection, etc.); population characteristics (e.g., age, sample size, nutritional information at baseline, breastfeeding status, etc.); intervention/comparison characteristics (e.g., exposure, tools used for measurement, etc.); outcomes (e.g., food frequency data, methods/tools used, age at outcome assessment, etc.); data analysis methods; control of confounding; funding obtained, and any conflict of interests.

Non-English language studies that were included in this review were translated from Portuguese (n=1) and Chinese (n=1) to English. Where any information was unclear, seemed incorrect, or missing, we contacted the authors for missing details (n=3).

Assessment of risk of bias in included studies

Quality assessments of included studies were conducted independently by four review authors (CO, KC, KR, LH). Any discrepancies between reviewers were resolved through discussion or by a fifth reviewer. Individual studies were critically appraised using the Cochrane Risk of Bias-2 (ROB-2) tool for randomized controlled trials and cluster-randomized controlled trials,[39] the NIH tool for observational cohort and cross-sectional studies,[40] and the ROBINS-I Tool for non-randomized studies.[41]

The Cochrane Risk of Bias-2 tool assesses RCT studies for risk of bias in the following domains: randomization process, deviations from the intended interventions, missing outcome data, outcome measurement, and the selection of the reported results. An overall risk of bias judgement was given to each study (low, high, some concerns).

Using the NIH tool, observational studies were rated good, fair, or poor, based on fourteen criteria covering the research question, participant population, analyses, timeframe, independent and dependent variables, attrition, and control of confounding variables.

The ROBINS-I tool assessed non-randomized trials in the following domains: bias due to confounding, bias in selection of study participants, bias in classification of interventions, bias due to deviations from intended interventions, bias due to missing data, bias in measurement of outcomes, and bias in selection of the reported results. Each study was given an overall risk of bias judgement (low, moderate, serious, critical). See Appendix B for further details about the quality assessment tools utilized in this review.

Grading Certainty of Evidence

The GRADE tool was used to assess the certainty of evidence for selected outcomes. The following outcomes were graded: wasting, underweight status, stunting, anemia, HAZ, % stability and change in vegetable and fruit consumption, linear growth, and ferritin <20µg/L.

The certainty of evidence was rated (very low, low, moderate, high) for each outcome in accordance with the GRADE framework. Evidence was downgraded based on five factors: risk of bias, inconsistency, indirectness, imprecision and publication bias. The criteria and reasons for downgrading were provided in explanatory footnotes in the GRADE tables. GRADE assessments are summarized in the Appendix C: Table C1: GRADE table for vegetables, Table C2: Grade table for fruit, Table C3 GRADE table for legumes and beans and Table C4 GRADE table for fruit and vegetables combined.

Data Synthesis

All experimental and observational study data were analyzed separately. Conducting meta-analysis was not possible for any outcome, given the heterogeneity across studies, including food group item, outcome metrics, and frequency categories. As such, a narrative synthesis was conducted for the FV and NPS food groups. Subgroup analyses according to age groups (6-8 months, 9-11 months and 12-23 months) was also not possible given the sparse and heterogeneous data.

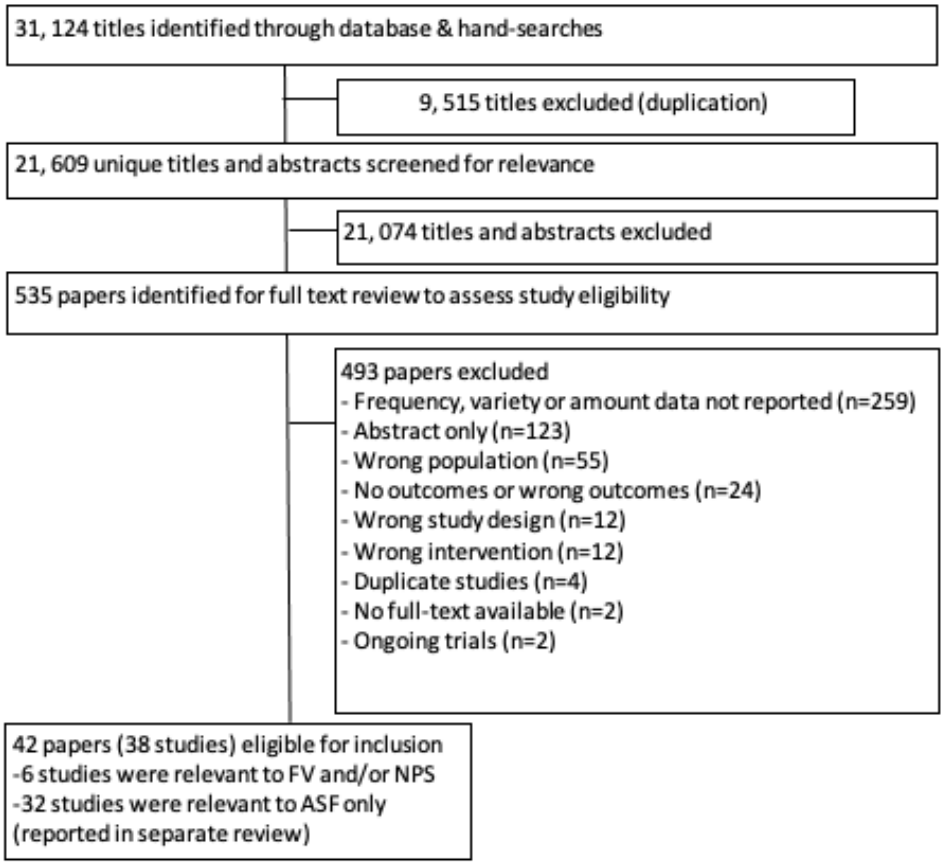
RESULTS

Description of Studies

Results of the Search

Our searches generated 31,124 records, of which 6 studies could not be retrieved from the CINAHL database. After removal of duplicates, we title/abstract screened 21,609 records, of which 535 studies were found to be potentially eligible for inclusion. Of these, 6 studies met our inclusion criteria for the FV/NPS review. See Figure 1 for study flow diagram.

Figure 1. Study flow diagram



Included Studies

We included six studies that analyzed data from 23,346 children from Senegal, Brazil, Indonesia, China, and Norway (n=2).[42-47]

See Appendix D for Study Demographics Table for further detail.

Study Design

Five of the included studies were cross-sectional studies [42-46] and one was a longitudinal cohort study.[47]

Location

Four studies were conducted in the following LMICs: Senegal, Brazil, Indonesia, and China.[42,44-46] Two studies were conducted in Norway.[43,47] One study was conducted in a rural location,[45] two

studies were conducted in urban locations,[43,46] two studies were conducted in a mix of urban and rural locations,[44,47] and one study did not report on setting.[42]

Participants

There were a total of 23,346 children in the included studies that ranged from 6 months to 7 years of age. Children were 6 to 23 months of age at the time of enrollment.

Description of Intervention

All six studies looked at frequency of consumption of vegetables.[42-47] Five studies looked at frequency of consumption of fruits.[42,44-47] One study looked at frequency of consumption of pulses[46] and one study looked at frequency of consumption of legumes.[42]

Frequency was defined differently in all 6 studies (for more detail see Appendix E for Detailed Study Characteristics of Included Studies Table). No studies reported on variety of FV or amount of NPS consumption.

In two of the studies, 24hr recalls were used to assess frequency of consumption of foods,[42,45] while another two used a food frequency questionnaire (FFQ) to assess frequency of consumption of foods over a week.[43,47] In one of the studies, both a 24hr recall and a FFQ that assessed food consumption over a week, were used.[46] In one study, no detail was provided on what tool was used to measure the exposure.[44]

Outcomes

<u>Wasting, Underweight, and Stunting</u>

One study reported wasting, underweight, and stunting prevalence based on body weight and recumbent length measurements.[42] Using the WHO definitions, underweight was defined as a weight-for-age z-score (WAZ) more than 2 standard deviation (SD) below the median (e.g., <-2 SD), wasting as a weight-for-length/height z-score (WLZ or WHZ) more than 2 SD below the median, and stunting as a length/height-for-age z-score (LAZ or HAZ) more than 2 SD below the median.

Another study reported the association between varying frequencies of FV consumption with stunting and underweight.[44] Stunting and underweight were defined based on WHO definitions, as described above.

<u>HAZ and Linear growth</u>

One study reported mean HAZ and linear growth based on recumbent height measurements.[45]

<u>Anemia</u>

One study reported anemia prevalence, defined as hemoglobin <11 g/dL,[46] WHO-defined hemoglobin cut-off for children 6 - 59 months.[48]

Iron Stores

One study reported on iron stores based on serum ferritin levels.[43]

Changes and stability in tracking FV intake

One study reported on stability and changes (increase and decrease) in FV consumption from 18 months to 36 months and 18 months to 7 years old in the same group of children.[47]

Excluded Studies

We excluded 493 studies. The most common reason for exclusion was that frequency, variety, or amount were not measured or reported in the data (n=259). 123 studies were excluded as they were abstracts only, 55 studies did not include our population of interest (e.g., children were unhealthy or not within the correct age range), 24 did not report outcomes or were not outcomes of interest, 12 studies provided an intervention that was not of interest, 12 were the wrong study design (e.g., qualitative), 4 turned out to be duplicates, and there were 2 ongoing trials.

Studies awaiting classification

We assessed two ongoing registered clinical trials as potentially eligible for inclusion (NCT02634749[49] and PACTR201906819960554[50]).

Two additional studies met our inclusion criteria, but we were unable to extract data in a usable form, as it was represented in graphs and could only be extracted as approximate values. Primary authors of both studies were contacted for data.[51,52]

Risk of Bias in Included Studies

Overall, we rated four of the included studies as having a high risk of bias/poor quality,[43-44,46-47] one study was rated as having a medium risk of bias/fair quality,[45] and one study was rated as having a low risk of bias/good quality[42] for internal validity based on the fourteen key concepts described below.

See Appendix F for risk of bias ratings for each included study.

Research question

In all six studies, the authors clearly described and defined their goal in conducting their research. Thus, there was a low risk of bias detected in all studies relating to the research question.[42-47]

Study population

In two of our included studies, the study population was clearly defined, and the participation rate of eligible persons was at least 50%.[42,45] These two studies had low risk of bias relating to the study population.

In one study, the study population was not specified or clearly defined, however the participation rate did appear to be at least 50% of eligible persons.[43] The remaining three studies both had unspecified study populations that were never clearly defined and did not have at least 50% of eligible persons participate, or were unclear.[44,46-47] For the three studies without 50% of eligible persons participating, there is increased risk of bias that the study population does not adequately represent the target population. Furthermore, for all four studies with unspecified target populations or study populations that were never clearly defined, it would be challenging to re-conduct these studies given the limited information on who to recruit, therefore we have scored them as having a high risk of bias.

Groups recruited from the same population and uniform eligibility criteria

For four of our included studies, the same eligibility criteria appeared to be applied for all subjects involved and the criteria included recruitment from the same population where applicable.[42-43,45-46] These four studies have a low risk of bias in this domain. For the remaining two studies, almost no detail is given on the study populations. For one study, the only detail provided on eligibility criteria was, “Any pregnant women recruited during the data collection timeframe”.[47] For the second study, there was no detail provided on eligibility criteria.[44] These two studies present a high risk of bias in this domain.

Sample size justification

Only one study provided a sample size justification,[42] while the remaining five did not.[43-47] However, given the observational design of these studies and their exploratory analyses, this was not considered a significant limitation or major source of bias.

Exposure assessed prior to outcome measurement

Given five of these studies are cross-sectional by design, the majority of our included studies did not assess the exposure prior to measurement of the outcome but rather this was done simultaneously.[42-46] Although this does not increase risk of bias, it does provide overall weaker evidence and for these five studies no causal relationship can be described between the exposure and outcome. The remaining study was a longitudinal cohort that did successfully

assess the exposure prior to measurement of the outcome,[47] resulting in a low risk of bias for this domain.

Sufficient timeframe to see an effect

Once again, given that the majority of our included studies are cross-sectional by design, these five studies did not allow time to see an effect but rather measured the exposure and outcome simultaneously.[42-46] The remaining study allotted a timeframe from 18 months to 7 years of age to detect an effect in changes and stability of FV consumption.[47] This is a sufficient and lengthy timeframe for an effect to be observed and thus presents low risk of bias in this domain.

Different levels of the exposure of interest

One of the eligibility criteria for this review was consumption of FV and/or NPS reported in terms of frequency, variety, or amount. Because of this, all six studies report different levels (different frequencies) of the exposure of interest.[42-47] Thus, there is low risk of bias for all included studies in this domain.

Exposure measures and assessment

For three of our included studies, the exposure measure was clearly defined, valid, reliable, and implemented consistently across all study participants using either a 24hr recall or a FFQ.[42-43,45] For these studies, there is low risk of bias in this domain. In two of the studies, both a FFQ and 24hr recall were used to assess the exposure, however, no details were provided on the validity or reliability of these tools.[46-47] The remaining study did not provide any detail on the exposure measures.[44] Thus, without detailed information on these tools it must be assumed that they are not valid or accurate, and so we conclude a high risk of bias for these three studies in this domain.

Repeated exposure assessment

For two of our included studies the exposure was assessed at three different time points.[42,47] This allowed for a more detailed look at changes in the exposure over time and more robust findings from the analyses conducted, providing a low risk of bias for these two studies in this domain. For four of the studies, the exposure was only assessed at one time point as these studies were cross-sectional by design.[43-46] Thus, this was not considered a significant limitation or major source of bias.

Outcome measures

For four of our included studies, the outcome measures were clearly defined, valid, reliable, and implemented consistently across all study participants.[42-43,45-46] For these studies, there is a low risk of bias in this domain. For the remaining two studies, there were limited details

provided on the outcome measure and it was challenging to determine whether there was validity.[44,47] Given the lack of detail provided, this study was rated as high risk in this domain.

Blinding of outcome assessors

In all six studies, no details are provided in the methodology of the paper in regard to potential blinding or masking of outcome assessors.[42-47] For this reason, it must be assumed that there is no blinding. This could be a major concern for those completing the data analysis portion of the study, as it is possible that bias could arise from knowing the exposure status of the study participant while looking for relationships with the outcome. Hence, these studies were rated as medium to high risk of bias in this domain.

Follow up rate

Bias arising due to loss to follow-up was not applicable for five of our included studies, given they were cross-sectional by design.[42-46] For the longitudinal cohort study, those lost to follow-up was significantly greater than 20%.[47] However, given the long timeframe of this study (5.5 years) this is to be expected. As the sample size at the final time point is still well over 1000 study participants, we characterized the risk of bias as low to medium for this study in this domain.

Statistical analyses

Two of our included studies reported that they accounted for confounding variables.[45-46] However, of the relevant data that we extracted, only one study actually adjusted its models for confounding variables.[45] This study was rated as low risk of bias for statistical analysis. The remaining four studies did not control for any confounding variables throughout their analyses.[42-44,47] Thus, these four studies, as well as the first study given that the data we extracted was not controlled for, were all rated as high risk of bias for statistical analysis.

Narrative Synthesis of Results

We included six studies that reported on the following outcomes: wasting, underweight, stunting, linear growth, HAZ, iron status, anemia, and changes and stability in FV intake later in life. All six of the included studies looked at the association between varying frequencies of FV or NPS consumption. None of our included studies looked at differing varieties or amounts of FV or NPS on dietary and health outcomes.

See Tables 3-6 for further detail below.

Table 3. Narrative Synthesis Table for Vegetables

Food Group	Study Design	Outcome	Study Name	No. Participants	Age range	Country	Narrative Synthesis	Other notes
V	CS	Wasting	Ahmad 2018[42]	392	6-23 months	Indonesia	Frequency of eating green leafy and orange vegetables was not associated with wasting. 20.5% (n=31), 25.2% (n=31), and 22% (n=26) of those who consumed green leafy and orange vegetables ≥4 times/week, 1-3 times/ week, and never were wasted, respectively (p=0.542 between groups).	
V	CS	Underweight	Ahmad 2018[42]	392	6-23 months	Indonesia	Frequency of eating green leafy and orange vegetables was not associated with being underweight. 25.8% (n=39), 26% (n=32), and 27.1% (n=32) of those who consumed green leafy and orange vegetables ≥4 times/week, 1-3 times/ week, and never were underweight, respectively (p=0.969 between groups).	
V	CS	Stunting	Ahmad 2018[42]	392	6-23 months	Indonesia	Frequency of eating green leafy and orange vegetables was not associated with stunting. 29.1% (n=44), 30.1% (n=37), and 23.7% (n=28) of those who consumed green leafy and orange vegetables ≥4 times/week, 1-3 times/ week, and never were stunted, respectively (p=0.491 between groups).	
V	CS	Height-for-age Z-score	Ntab 2005[45]	165	6-42 months	Senegal	In breastfed children aged 9-23 months, those who consumed vegetables/ leaves 0-2 days/ week and ≥ 3 days/week had a mean HAZ of -1.01 (SD=0.93) and -0.58 (SD=0.92), respectively. In age-adjusted models, mean HAZ was -1.01 (p=0.052) and -0.59 (p<0.06), respectively, demonstrating a trend that points to lower HAZ with less vegetable consumption.	Adjusted model by child age.

V	CS	Linear Growth	Ntab 2005[45]	165	6-42 months	Senegal	Frequent consumption of vegetables had an inverse relationship to linear growth (means: 8.3cm and 7.4cm height increment over the preceding 7 months for rare and frequent consumption, respectively, p=0.041).	Adjusted model by child age, sex, malaria study intervention group, maternal height, BMI, schooling, and number of children 5 y old.
V	CS	Ferritin >20µg/L	Wandel 1996[43]	90	1 year old	Norway	Those who consumed vegetables once/day, compared to <once/day, were significantly more likely to have low iron stores (ferritin values <20µg/L) (Reg coefficient= -2.7, p=0.02). Although the results suggest a negative effect of feeding vegetables once/day, findings were inconsistent as this relationship was not seen for those eating vegetables more frequently (several times/day) compared to <once/day. Additionally, there was no significant relationship observed between the feeding frequency of vegetables and the likelihood of having very low iron stores (ferritin <15µg/L).	
V	CS	Anemia	Silva 2007[46]	205	6-12 months	Brazil	Consumption of dark green vegetables was not associated with anemia prevalence. For those who did not consume dark green vegetables in the past 24 hours, compared to those who did, the odds of having anemia were 1.21 times greater (95% CI: 0.67-2.21, p=0.502). 55.6% (n=65) and 60.2% (n=53) of those who consumed dark green vegetables in the past 24 hours did not have anemia, respectively.	Anemia= Hemoglobin <11 g/dL. Exclusive breastfeeding time (Days) ≥60: 62 (anemia); 52 (non-anemia)

								<60: 56 (anemia); 35 (non-anemia)
V	Cohort	% Stability and Change in Vegetable Consumption	Bjelland 2013[47]	1,783	18 months to 7 years old	Norway	From 18 months to 36 months of age, 0% of boys decreased, 75.3% remained stable, and 24.7% increased their vegetable consumption from low (≤ 5 times/week). From 18 months to 7 years of age, 0% of boys decreased, 63% remained stable, and 37% increased their vegetable consumption from low (≤ 5 times/week).	Overall vegetable consumption at 18 months was positively associated with overall vegetable consumption at 36 months (spearman's rho=0.36) and at 7 years of age (spearman's rho=0.28).
V	Cohort	% Stability and Change in Vegetable Consumption	Bjelland 2013[47]	1,075	18 months to 7 years old	Norway	From 18 months to 36 months of age, 55.1% of boys decreased, 32.7% remained stable, and 12.3% increased their vegetable consumption from medium (5.1-7 times/week). From 18 months to 7 years of age, 42.6% of boys decreased, 21.2% remained stable, and 36.2% increased their vegetable consumption from medium (5.1-7 times/week).	
V	Cohort	% Stability and Change in Vegetable Consumption	Bjelland 2013[47]	1,649	18 months to 7 years old	Norway	From 18 months to 36 months of age, 74% of boys decreased, 26% remained stable, and 0% increased their vegetable consumption from high (>7 times/week). From 18 months to 7 years of age, 51.2% of boys decreased, 48.8% remained stable, and 0% increased their vegetable consumption from high (>7 times/week).	
								Overall vegetable

V	Cohort	% Stability and Change in Vegetable Consumption	Bjelland 2013[47]	1,677	18 months to 7 years old	Norway	From 18 months to 36 months of age, 0% of girls decreased, 73.3% remained stable, and 26.7% increased their vegetable consumption from low (≤ 5 times/week). From 18 months to 7 years of age, 0% of girls decreased,59.8 % remained stable, and 40.2% increased their vegetable consumption from low (≤ 5 times/week).	consumption at 18 months was positively associated with overall vegetable consumption at 36 months (spearman's rho=0.37) and at 7 years of age (spearman's rho=0.31).
V	Cohort	% Stability and Change in Vegetable Consumption	Bjelland 2013[47]	1,012	18 months to 7 years old	Norway	From 18 months to 36 months of age, 54.2% of girls decreased, 34.3% remained stable, and 11.5% increased their vegetable consumption from medium (5.1-7 times/week). From 18 months to 7 years of age, 40.5% of girls decreased, 19.5% remained stable, and 40% increased their vegetable consumption from medium (5.1-7 times/week).	
V	Cohort	% Stability and Change in Vegetable Consumption	Bjelland 2013[47]	1,600	18 months to 7 years old	Norway	From 18 months to 36 months of age, 72.3% of girls decreased, 27.7% remained stable, and 0% increased their vegetable consumption from high (>7 times/week). From 18 months to 7 years of age, 43.9% of girls decreased, 56.1% remained stable, and 0% increased their vegetable consumption from high (>7 times/week).	

Table 4. Narrative Synthesis Table for Fruits

Food Group	Study Design	Outcome	Study Name	No. Participants	Age range	Country	Narrative Synthesis	Other notes
F	CS	Wasting	Ahmad 2018[42]	392	6-23 months	Indonesia	Frequency of eating fruits was not associated with wasting. 19.9% (n=28), 26.1% (n=42), and 20% (n=18) of those who consumed fruits ≥3 times/week, 1-2 times/ week, and never were wasted, respectively (p=0.356 between groups).	
F	CS	Underweight	Ahmad 2018[42]	392	6-23 months	Indonesia	Frequency of eating fruits was not associated with being underweight. 26.2% (n=37), 26.1% (n=42), and 26.7% (n=24) of those who consumed fruits ≥3 times/week, 1-2 times/ week, and never were underweight, respectively (p=0.995 between groups).	
F	CS	Stunting	Ahmad 2018[42]	392	6-23 months	Indonesia	Frequency of eating fruits was not associated with stunting. 30.5% (n=43), 25.5% (n=41), and 27.8% (n=25) of those who consumed fruits ≥3 times/week, 1-2 times/ week, and never were stunted, respectively (p=0.623 between groups).	
F	CS	Height-for-age Z-score	Ntab 2005[45]	165	6-42 months	Senegal	In breastfed children aged 9-23 months, those who consumed fruit 0-2 days/ week and ≥ 3 days/week had a mean HAZ of -1.02 (SD=0.93) and -0.75 (SD=0.94), respectively. In age-adjusted models, mean HAZ was -1.04 (p=0.051) and -0.71 (p=0.059), respectively, demonstrating a trend that points to lower HAZ with less fruit consumption.	Adjusted model by child age.
F	CS	Linear Growth	Ntab 2005[45]	165	6-42 months	Senegal	Fruit consumption was positively associated with linear growth in fully adjusted models (means: 7.9cm and 8.7cm height increment over the preceding 7 months for rare and frequent consumption, respectively, p=0.027).	Adjusted model by child age, sex, malaria study intervention group, maternal

								height, BMI, schooling, and number of children 5 y old.
F	CS	Anemia	Silva 2007[46]	205	6-12 months	Brazil	Fruit consumption was not significantly associated with anemia prevalence when considering consumption in the past 24 hours. For those who did not consume fruit in the past 24 hours, compared to those who did, the odds of having anemia were 1.24 times greater (95% CI: 0.25-6.75, p=0.537). 57.4% (n=113) and 62.5% (n=5) of those who consumed fruit in the past 24 hours and not, had anemia, respectively. When looking at daily vs < than daily frequencies, fruit consumption was significantly associated with anemia prevalence. For those who consumed fruit < daily, compared to daily, the odds were 2.29 times greater (1.29-4.06, p=0.004) of having anemia. In the adjusted model, those who consumed fruit < daily, compared to daily, had 1.88 (95% CI:1.03-3.42, p=0.003) greater odds of having anemia. 48.7% (n=55) and 68.5% (n=63) of those who consumed fruit daily and <daily had anemia, respectively.	Anemia= Hemoglobin <11 g/dL. Adjusted model by family income per capita and consumption of medicated iron supplements. Exclusive breastfeeding time (Days) ≥60: 62 (anemia); 52 (non-anemia) <60: 56 (anemia); 35 (non-anemia).
F	Cohort	% Stability and Change in Fruit Consumption	Bjelland 2013[47]	1,659	18 months to 7 years old	Norway	From 18 months to 36 months of age, 0% of boys decreased, 37.6% remained stable, and 62.4% increased their fruit consumption from low (≤ 5 times/week). From 18 months to 7 years of age, 0% of boys decreased, 42.2% remained stable, and 57.8% increased their fruit consumption from low (≤ 5 times/week).	Overall fruit consumption at 18 months was positively associated with overall fruit consumption at 36 months (spearman's

								rho=0.36) and at 7 years of age (spearman's rho=0.23).
F	Cohort	% Stability and Change in Fruit Consumption	Bjelland 2013[47]	2,535	18 months to 7 years old	Norway	From 18 months to 36 months of age, 12.8% of boys decreased, 39% remained stable, and 48.2% increased their fruit consumption from medium (5.1-13.9 times/week). From 18 months to 7 years of age, 24.7% of boys decreased, 59.8% remained stable, and 15.5% increased their vegetable consumption from medium (5.1-13.9 times/week).	
F	Cohort	% Stability and Change in Fruit Consumption	Bjelland 2013[47]	290	18 months to 7 years old	Norway	From 18 months to 36 months of age, 24.8% of boys decreased, 75.2% remained stable, and 0% increased their fruit consumption from high (>14 times/week). From 18 months to 7 years of age, 67.9% of boys decreased, 32.1% remained stable, and 0% increased their fruit consumption from high (>14 times/week).	
F	Cohort	% Stability and Change in Fruit Consumption	Bjelland 2013[47]	1,601	18 months to 7 years old	Norway	From 18 months to 36 months of age, 0% of girls decreased, 36.3% remained stable, and 63.7% increased their fruit consumption from low (≤ 5 times/week). From 18 months to 7 years of age, 0% of girls decreased, 38.4% remained stable, and 61.6% increased their fruit consumption from low (≤ 5 times/week).	Overall fruit consumption at 18 months was positively associated with overall fruit consumption at 36 months (spearman's rho=0.36) and at 7 years of age (spearman's rho=0.24).
F	Cohort	% Stability and Change	Bjelland 2013[47]	2,375			From 18 months to 36 months of age, 11.8% of girls decreased, 41.3% remained	

		in Fruit Consumption			18 months to 7 years old	Norway	stable, and 46.9% increased their fruit consumption from medium (5.1- 13.9 times/week). From 18 months to 7 years of age, 20.3% of girls decreased, 62.3% remained stable, and 17.4% increased their fruit consumption from medium (5.1-13.9 times/week).	
F	Cohort	% Stability and Change in Fruit Consumption	Bjelland 2013[47]	271	18 months to 7 years old	Norway	From 18 months to 36 months of age, 23.6% of girls decreased, 76.4% remained stable, and 0% increased their fruit consumption from high (>14 times/week). From 18 months to 7 years of age, 66.8% of girls decreased, 33.2% remained stable, and 0% increased their fruit consumption from high (>14 times/week).	

Table 5. Narrative Synthesis Table for Legumes and Beans

Food Group	Study Design	Outcome	Study Name	No. Participants	Age range	Country	Narrative Synthesis	Other notes
Legumes	CS	Wasting	Ahmad 2018[42]	392	6-23 months	Indonesia	Frequency of eating legumes was not associated with wasting. 25.9% (n=15), 24% (n=37), and 20% (n=36) of those who consumed legumes ≥3 times/week, 1-2 times/ week, and never were wasted, respectively (p=0.542 between groups).	
Legumes	CS	Underweight	Ahmad 2018[42]	392	6-23 months	Indonesia	Frequency of eating legumes was not associated with being underweight. 36.2% (n=21), 25.3% (n=39), and 23.9% (n=43) of those who consumed legumes ≥3 times/week, 1-2 times/ week, and never were underweight, respectively (p=0.174 between groups).	
Legumes	CS	Stunting	Ahmad 2018[42]	392	6-23 months	Indonesia	Frequency of eating legumes was not associated with stunting. 32.8% (n=19), 27.9% (n=43), and 26.1% (n=47) of those who consumed legumes ≥3 times/week,	

							1-2 times/ week, and never were stunted, respectively (p=0.618 between groups).	
Beans	CS	Anemia	Silva 2007[46]	205	6-12 months	Brazil	Consumption of beans was not significantly associated with anemia prevalence. For those who consumed beans < daily, compared to daily, the odds of having anemia were 0.8 times less (95% CI: 0.36-1.78, p=0.550). 58.5% (n=100) and 52.9% (n=18) of those who consumed beans daily and <daily had anemia, respectively.	Anemia= Hemoglobin <11 g/dL. Exclusive breastfeeding time (Days) ≥60: 62 (anemia); 52 (non-anemia) <60: 56 (anemia); 35 (non-anemia)

Table 6. Narrative Synthesis for FV combined

Food Group	Study Design	Outcome	Study Name	No. Participants	Age range	Country	Narrative Synthesis	Other notes
F + V	CS	Stunting	Chang 2007[44]	13,107	6-24 months	China	Frequency of eating vegetables and fruits was associated with stunting prevalence. For those who consumed vegetables and fruits weekly, monthly, and < once per month or none, compared to daily, the odds of being stunted were 1.739 (p=0.00), 1.698 (p=0.03), and 1.768 (p=0.00) times greater, respectively.	
F + V	CS	Underweight	Chang 2007[44]	13,107	6-24 months	China	Frequency of eating vegetables and fruits was associated with underweight prevalence. For those who consumed vegetables and fruits weekly, monthly, and < once per month or none, compared to daily, the odds of being underweight were 1.908 (p=0.00), 1.566 (p=0.10), and 1.478 (p=0.01) times greater, respectively.	

Wasting, Underweight, and Stunting

Two studies reported on anthropometric outcomes based on varying frequencies of FV or NPS consumption. However, given the high heterogeneity of the two studies (different food group items and frequency categories), we were unable to pool the data and instead a narrative synthesis is described below.

One study with a total of 392 children aged 6-23 months evaluated the association of fruit, vegetable, and legume consumption at different frequencies with wasting, underweight, and stunting prevalence.[42]

This study did not find an association between less frequent consumption of green leafy and orange vegetables and wasting, underweight, or stunting among children. Frequency was categorized as ≥4 times/week, 1-3 times/week, and never. 20.5% (n=31), 25.2% (n=31), and 22% (n=26) of those who consumed green leafy and orange vegetables ≥4 times/week, 1-3 times/ week, and never were wasted, respectively (p=0.542 between groups). 25.8% (n=39), 26% (n=32), and 27.1% (n=32) of those who consumed green leafy and orange vegetables ≥4 times/week, 1-3 times/ week, and never were underweight, respectively (p=0.969 between groups). 29.1% (n=44), 30.1% (n=37), and 23.7% (n=28) of those who consumed green leafy and orange vegetables ≥4 times/week, 1-3 times/ week, and never were stunted, respectively (p=0.491 between groups).

This study also did not find an association between less frequent consumption of fruits and wasting, underweight, or stunting. Frequency was categorized as ≥3 times/week, 1-2 times/ week, and never. 19.9% (n=28), 26.1% (n=42), and 20% (n=18) of those who consumed fruits ≥3 times/week, 1-2 times/ week, and never were wasted, respectively (p=0.356 between groups). 26.2% (n=37), 26.1% (n=42), and 26.7% (n=24) of those who consumed fruits ≥3 times/week, 1-2 times/ week, and never were underweight, respectively (p=0.995 between groups). 30.5% (n=43), 25.5% (n=41), and 27.8% (n=25) of those who consumed fruits ≥3 times/week, 1-2 times/ week, and never were stunted, respectively (p=0.623 between groups).

Lastly, this study did not find an association between less frequent consumption of legumes and wasting, underweight, or stunting. Frequency was defined as ≥3 times/week, 1-2 times/week, and never. 25.9% (n=15), 24% (n=37), and 20% (n=36) of those who consumed legumes ≥3 times/week, 1-2 times/ week, and never were wasted, respectively (p=0.542 between groups). 36.2% (n=21), 25.3% (n=39), and 23.9% (n=43) of those who consumed legumes ≥3 times/week, 1-2 times/week, and never were underweight, respectively (p=0.174 between groups). 32.8% (n=19), 27.9% (n=43), and 26.1% (n=47) of those who consumed legumes ≥3 times/week, 1-2 times/ week, and never were stunted, respectively (p=0.618 between groups).

We were unable to present these results in a forest plot as only p-values were reported in the paper.

A second study, with a total of 13,107 children aged 6-24 months evaluated the association of FV consumption at different frequencies with stunting and underweight prevalence.[44]

This study found an association between less frequent consumption of FV and stunting and underweight prevalence among children. Frequency was categorized as daily, weekly, monthly, and less than once per month or none. For those who consumed vegetables and fruits weekly, monthly, and less than once per month or none, compared to daily, the odds of being stunted were 1.739 (p=0.00), 1.698 (p=0.03), and 1.768 (p=0.00) times greater, respectively. Similarly, for those who consumed vegetables and fruits weekly, monthly, and less than once per month or none, compared to daily, the odds of being underweight were 1.908 (p=0.00), 1.566 (p=0.10), and 1.478 (p=0.01) times greater, respectively.

Figure 2. Forest plot demonstrating association between weekly vs daily FV consumption and stunting prevalence

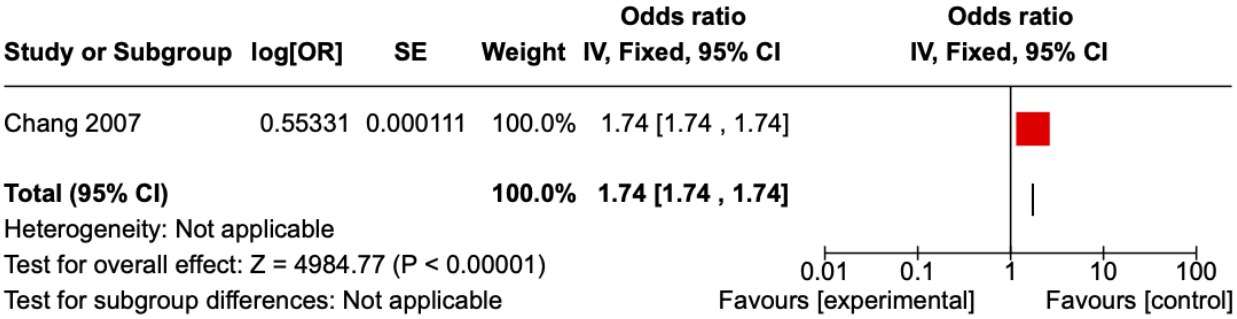


Figure 3. Forest plot demonstrating association between monthly vs daily FV consumption and stunting prevalence

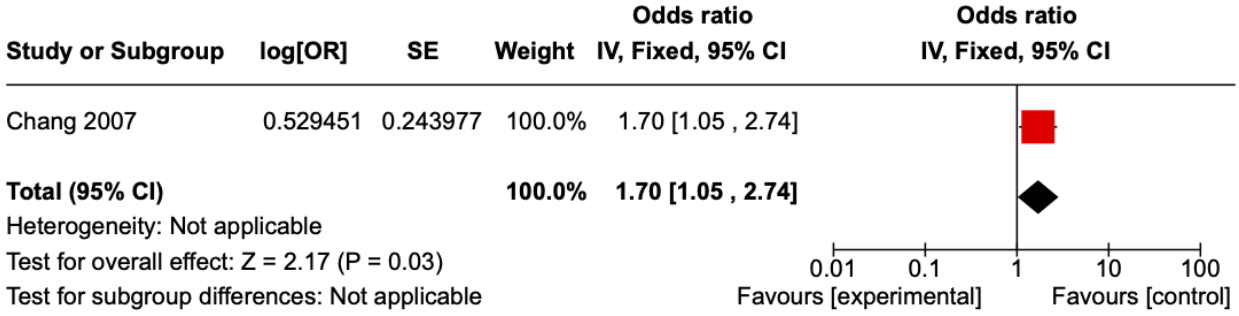


Figure 4. Forest plot demonstrating association between < than once per month or no vs daily FV consumption and stunting prevalence

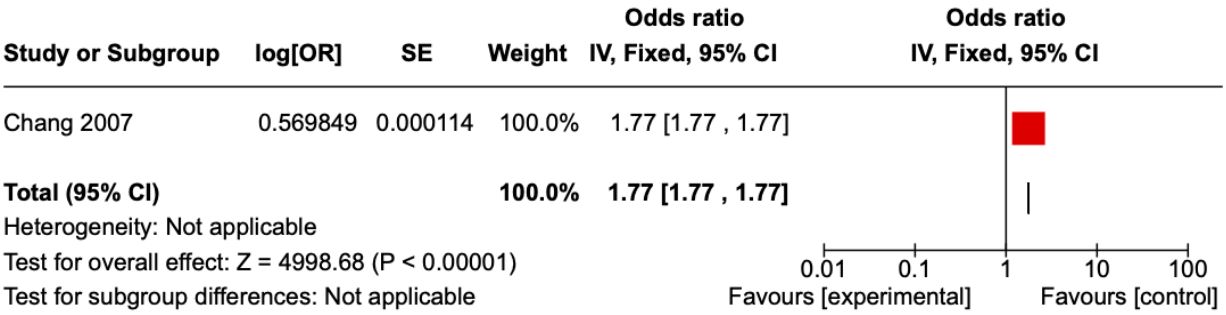


Figure 5. Forest plot demonstrating association between weekly vs daily FV consumption and underweight prevalence

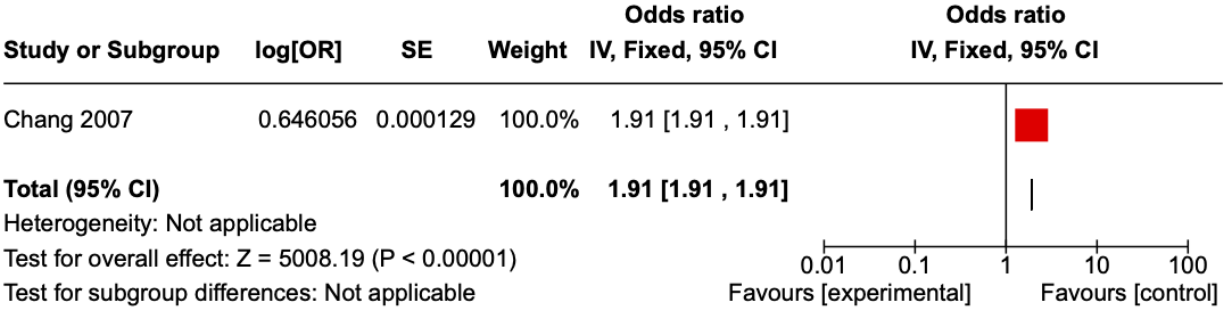


Figure 6. Forest plot demonstrating association between monthly vs daily FV consumption and underweight prevalence

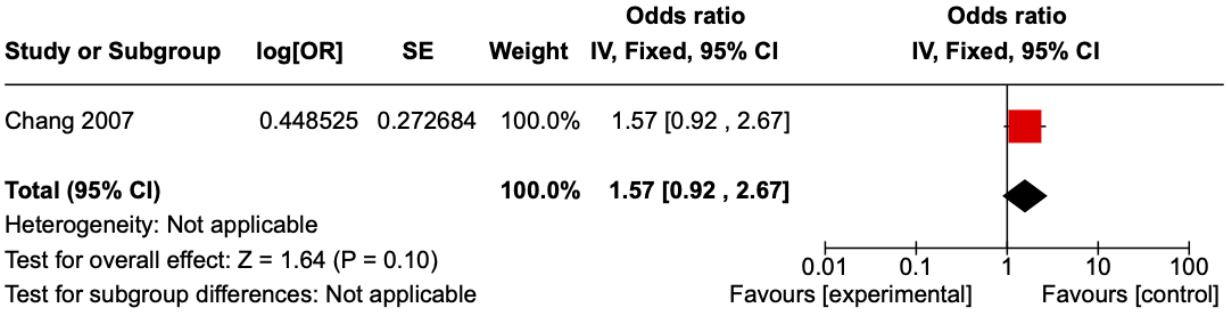
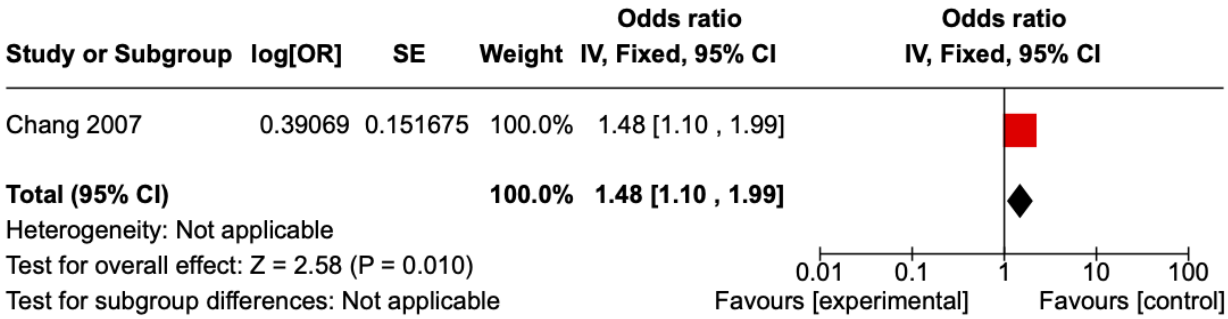


Figure 7. Forest plot demonstrating association between < than once per month or no vs daily FV consumption and underweight prevalence



HAZ and Linear growth

One study with a total of 165 children aged 9-23 months evaluated the association of frequent versus infrequent fruit and vegetable consumption with HAZ and linear growth based on recumbent height measurements.[45]

For the outcome of HAZ, frequency was defined as 0-2 days/week and ≥ 3 days/week. In breastfed children aged 9-23 months, those who consumed vegetables/leaves 0-2 days/week and ≥ 3 days/week had a mean HAZ of -1.01 (SD=0.93) and -0.58 (SD=0.92), respectively. In age-adjusted models, mean HAZ was -1.01 (p=0.052) and -0.59 (p<0.06), respectively (SD was not reported for any adjusted models). In breastfed children aged 9-23 months, those who consumed fruit 0-2 days/week and ≥ 3 days/week had a mean HAZ of -1.02 (SD=0.93) and -0.75 (SD=0.94), respectively. In age-adjusted models, mean HAZ was -1.04 (p=0.051) and -0.71 (p=0.059), respectively. Based on these results for both vegetable and fruit consumption, there is a trend that points to lower HAZ among children who consume vegetables and fruits less frequently.

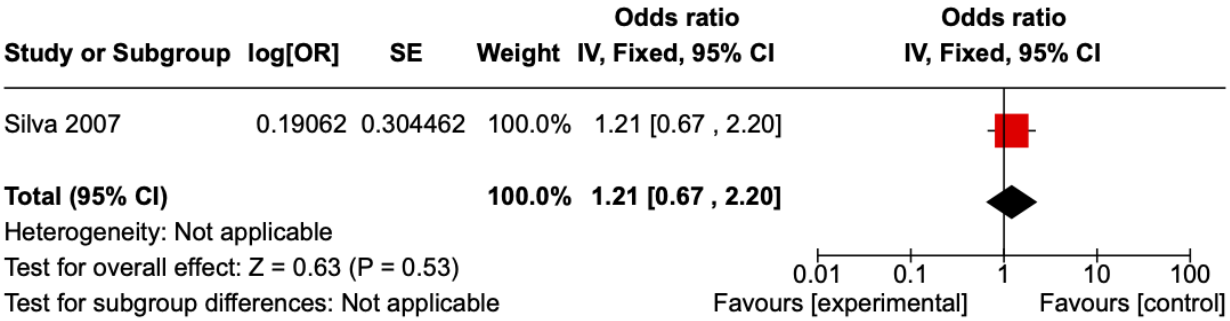
For linear growth, frequency was defined as rare or frequent consumption; however, no further details were provided on a more specific definition of what was considered rare or frequent consumption. Among children who consumed fruit rarely versus frequently over the preceding 7 months, their linear growth increased by a mean of 8.3cm versus 7.4cm, respectively (p=0.041). However, in the fully adjusted model, fruit consumption was positively associated with linear growth, with means of 7.9cm and 8.7cm height increments over the preceding 7 months among children who consumed fruit rarely and frequently, respectively (p=0.027). The model was adjusted for child age, sex, malaria study intervention group (this study was nested into the Intermittent Preventive malaria Treatment in children (IPTc) study, where children were randomized into receiving artesunate plus sulfadoxine pyromethamine or a placebo), maternal height, BMI, schooling, and number of children 5yrs old in the household.

Anemia

One study with a total of 205 children aged 6-12 months evaluated the association of vegetable, fruit, and bean consumption with anemia prevalence. Anemia was defined as hemoglobin <11 g/dL.[46]

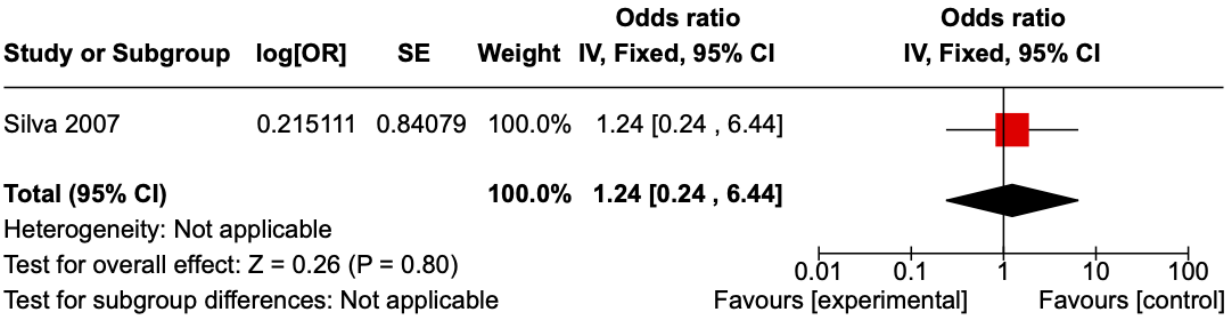
For vegetables, frequency was defined as consumption of the food item in the previous 24 hours versus no consumption in the previous 24 hours. 55.6% (n=65) and 60.2% (n=53) of those who consumed dark green vegetables in the past 24 hours and did not consume dark green vegetables in the past 24 hours had anemia, respectively. Authors found that lack of consumption of dark green vegetables was not associated with higher anemia prevalence. For those who did not consume dark green vegetables in the past 24 hours, compared to those who did, the odds of being anemic were 1.21 times greater (95% CI: 0.67-2.21, p=0.502), though results were not statistically significant.

Figure 8. Forest plot demonstrating association between dark green vegetable consumption in past 24 hours vs not and anemia prevalence



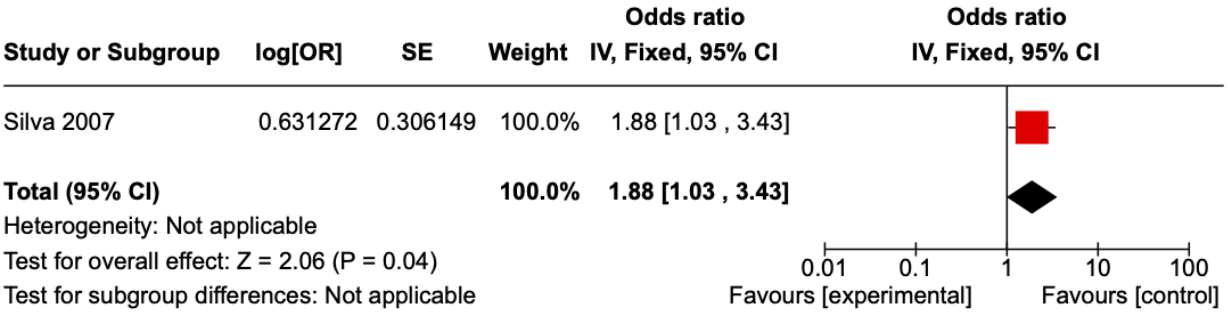
For fruit, frequency was defined in one model as consumption of the food item in the previous 24 hours versus no consumption in the previous 24 hours. Additionally, a second model defined frequency as consumption of the food item daily versus less than daily. When looking at the metric of past 24 hours, 57.4% (n=113) of all children who consumed fruit were anemic and 62.5% (n=5) of those who did not consume fruit in the past 24 hours were anemic. Furthermore, lack of fruit consumption was not significantly associated with higher anemia prevalence. For those who did not consume fruit in the past 24 hours, compared to those who did, the odds of being anemic were 1.24 times greater (95% CI: 0.25-6.75, p=0.537) but did not reach statistical significance.

Figure 9. Forest plot demonstrating association between fruit consumption in past 24 hours vs not and anemia prevalence



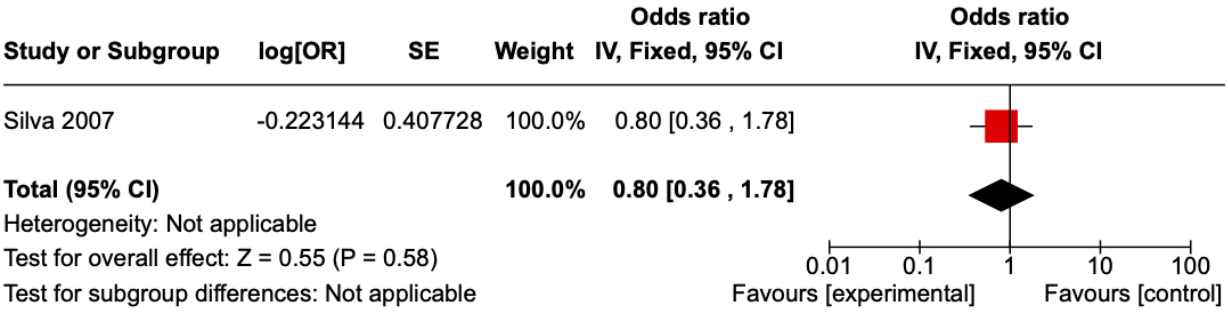
When looking at the metric of daily versus less than daily consumption of fruits and anemia prevalence, 48.7% (n=55) of all children who consumed fruit daily were anemic and 68.5% (n=63) of those who did not consume fruit daily were anemic. Conversely to the 24 hour metric, there was a statistically significant association when using the daily versus less than daily metric. For those who consumed fruit less than daily, compared to daily, the odds of being anemic were 2.29 times greater (95% CI: 1.29-4.06, p=0.004). When adjusted, those who consumed fruit less than daily, compared to daily, had 1.88 times greater odds (95% CI: 1.03-3.42, p=0.003) of being anemic. The model was adjusted by family income per capita and for consumption of iron supplements.

Figure 10. Forest plot demonstrating association between < daily vs daily fruit consumption and anemia prevalence



For beans, frequency was defined as consumption of the food item daily versus less than daily. 58.5% (n=100) and 52.9% (n=18) of those who consumed beans daily and less than daily had anemia, respectively. Lower consumption of beans was not significantly associated with a higher anemia prevalence. For those who consumed beans less than daily, compared to daily, the odds of being anemic were 0.8 times less (95% CI: 0.36-1.78, p=0.550).

Figure 11. Forest plot demonstrating association between < daily vs daily bean consumption and anemia prevalence

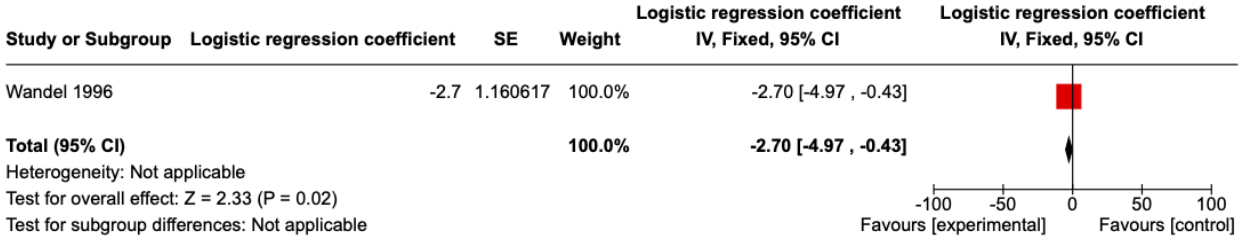


Iron Stores

One study with a total of 90 children aged 12 months old evaluated the association of vegetable consumption with iron stores based on serum ferritin levels.[43] Frequency was defined as consumption several times per day, once per day, and less than once per day.

Those who consumed vegetables once per day, compared to less than once per day, were significantly more likely to have ferritin values below the cut-off point of 20µg/L (Reg coefficient= -2.7, p=0.02). However, although the results suggest a negative association with feeding vegetables once per day, the findings were inconsistent given that this relationship was not seen for those eating vegetables more frequently (several times per day) compared to less than once per day. Additionally, there was no significant relationship observed between the feeding frequency of vegetables and the likelihood of having very low iron stores (ferritin 15µg/L).

Figure 12. Forest plot demonstrating association between vegetable consumption once per day vs < once per day and low iron stores (ferritin <20µg/L)



Changes and stability in tracking FV intake

One cohort study with a total of 9,490 children aged 18 months at baseline evaluated the association between vegetable and fruit consumption with stability and changes (increase and decrease) in vegetable and fruit consumption at 36 months and 7 years of age.[47] Frequency was categorized into 3 levels: low, medium, and high. For vegetables, low consumption was defined as ≤ 5 times/week, medium was defined as 5.1-7 times/week, and high was defined as >7 times/week. For fruits, low consumption was defined as ≤ 5 times/week, medium was defined as 5.1-13.9 times/week, and high was defined as >14 times/week.

<u>Vegetables</u>

In boys from 18 months to 36 months of age, 75.3% remained stable and 24.7% increased their vegetable consumption from low (≤ 5 times/week). From 18 months to 36 months of age, 55.1% of boys decreased, 32.7% remained stable, and 12.3% increased their vegetable consumption from medium (5.1-7 times/week). From 18 months to 36 months of age, 74% of boys decreased and 26% remained stable in their vegetable consumption from high (>7 times/week). Overall vegetable consumption at 18 months was positively associated with overall vegetable consumption at 36 months (spearman's rho=0.36).

In boys from 18 months to 7 years of age, 63% remained stable and 37% increased their vegetable consumption from low (≤ 5 times/week). From 18 months to 7 years of age, 42.6% of boys decreased, 21.2% remained stable, and 36.2% increased their vegetable consumption from medium (5.1-7 times/week). From 18 months to 7 years of age, 51.2% of boys decreased and 48.8% remained stable in their vegetable consumption from high (>7 times/week). Overall vegetable consumption at 18 months was positively associated with overall vegetable consumption at 7 years of age (spearman's rho=0.28).

In girls from 18 months to 36 months of age, 73.3% remained stable and 26.7% increased their vegetable consumption from low (≤ 5 times/week). From 18 months to 36 months of age, 54.2% of girls decreased, 34.3% remained stable, and 11.5% increased their vegetable consumption from medium (5.1-7 times/week). From 18 months to 36 months of age, 72.3% of girls decreased, and 27.7% remained stable in their vegetable consumption from high (>7 times/week). Overall vegetable consumption at 18 months was positively associated with overall vegetable consumption at 36 months (spearman's rho=0.37).

In girls from 18 months to 7 years of age, 59.8 % remained stable and 40.2% increased their vegetable consumption from low (≤ 5 times/week). From 18 months to 7 years of age, 40.5% of girls decreased, 19.5% remained stable, and 40% increased their vegetable consumption from medium (5.1-7 times/week). From 18 months to 7 years of age, 43.9% of girls decreased and 56.1% remained stable in their vegetable consumption from high (>7 times/week). Overall vegetable consumption at 18 months was positively associated with overall vegetable consumption at 7 years of age (spearman's rho=0.31).

Fruit

In boys from 18 months to 36 months of age, 37.6% remained stable and 62.4% increased their fruit consumption from low (≤ 5 times/week). From 18 months to 36 months of age, 12.8% of boys decreased, 39% remained stable, and 48.2% increased their fruit consumption from medium (5.1-13.9 times/week). From 18 months to 36 months of age, 24.8% of boys decreased and 75.2% remained stable in their fruit consumption from high (>14 times/week). Overall fruit consumption at 18 months was positively associated with overall fruit consumption at 36 months (spearman's rho=0.36).

In boys from 18 months to 7 years of age, 42.2% remained stable and 57.8% increased their fruit consumption from low (≤ 5 times/week). From 18 months to 7 years of age, 24.7% of boys decreased, 59.8% remained stable, and 15.5% increased their vegetable consumption from medium (5.1-13.9 times/week). From 18 months to 7 years of age, 67.9% of boys decreased and 32.1% remained stable in their fruit consumption from high (>14 times/week). Overall fruit consumption at 18 months was positively associated with overall fruit consumption at 7 years of age (spearman's rho=0.23).

In girls from 18 months to 36 months of age, 36.3% remained stable and 63.7% increased their fruit consumption from low (≤ 5 times/week). From 18 months to 36 months of age, 11.8% of

girls decreased, 41.3% remained stable, and 46.9% increased their fruit consumption from medium (5.1- 13.9 times/week). From 18 months to 36 months of age, 23.6% of girls decreased and 76.4% remained stable in their fruit consumption from high (>14 times/week). Overall fruit consumption at 18 months was positively associated with overall fruit consumption at 36 months (spearman's rho=0.36).

In girls from 18 months to 7 years of age, 38.4% remained stable and 61.6% increased their fruit consumption from low (≤ 5 times/week). From 18 months to 7 years of age, 20.3% of girls decreased, 62.3% remained stable, and 17.4% increased their fruit consumption from medium (5.1-13.9 times/week). From 18 months to 7 years of age, 66.8% of girls decreased and 33.2% remained stable in their fruit consumption from high (>14 times/week). Overall fruit consumption at 18 months was positively associated with overall fruit consumption at 36 months (spearman's rho= 0.24).

DISCUSSION

Summary of Main Results

Our review aimed to assess the effects of differing frequencies and varieties of FV and differing frequencies and amounts of NPS on dietary and health outcomes, though available data did not allow us to establish causality. We found six relevant studies, five of which were cross-sectional studies and one of which was a longitudinal cohort study. The longitudinal cohort study duration was 5.5 years, with children ranging in age from 18 months to 7 years. Two studies assessed the same outcome of stunting and underweight,[42,44] however there was high methodological and clinical heterogeneity between the two studies and so we were unable to pool results. The remaining four studies all assessed different outcomes, and frequencies of food consumption were reported differently throughout. Outcomes reported included: wasting, underweight, stunting, linear growth, HAZ, iron status, anemia, and changes and stability in FV intake later in life. For studies that reported on vegetable consumption, three of the studies did not describe further what type of vegetables were consumed, one study grouped green leafy and orange vegetables together, another study grouped vegetables and leaves together, and the final study described dark green vegetables. For studies that reported on fruit, no study described in more detail what fruits were consumed. For studies that reported on NPS, one study reported legume consumption and one study reported bean consumption. Of important note, only two studies adjusted models in their analyses for confounding variables, which needs to be considered when discussing the following results. The total sample size was 23,346 between the six studies of children ranging from 6 to 23 months of age at time of enrollment.

It is uncertain whether varying frequencies of consumption of dark green leafy and orange vegetables, fruits, or legumes reduces the prevalence of wasting, underweight, or stunting because the certainty of this evidence is very low.[42] This study reported no association between varying frequencies of consumption of dark green leafy and orange vegetables, fruits, or

legumes and wasting, underweight, or stunting. These findings were contrary to our expectation, as we would expect different frequencies of FV and NPS to correlate with nutritional status, and may be reflective of the limited number of studies that were identified. Given that these foods provide important sources of vitamins and minerals, one would assume that consuming them in greater frequencies would be associated with improved micronutrient (and protein, in the case of legumes) status which could translate into a decrease in wasting, underweight, and stunting prevalence. While Ahmad[42] did not demonstrate an association, Chang 2007[44] did find that less frequent consumption of FV was associated with a greater prevalence of stunting and underweight. However, once again, the certainty of this evidence was very low and thus we cannot determine if this result is close to the truth or not. To our knowledge, no other study to date has looked specifically at frequency of FV or NPS consumption and its association with nutritional status in infants. Although none of our included studies reported on differing varieties of FV or amounts of NPS, previous research in this area shows that complementary feeding practices (e.g., increased variety of FV, NPS, and other important food groups) are significantly correlated with nutritional status in infants.[53-55] These studies, though answering a similar research question, could not be included in our review because of lack of extractable data for outcomes or non-eligible exposure variables.

It is uncertain whether varying frequencies of consumption of vegetables/leaves or fruit improves HAZ or linear growth because the certainty of this evidence is very low.[45] This study demonstrated that less frequent consumption of vegetables/leaves and fruit were both associated with lower HAZ and more frequent fruit consumption was positively associated with linear growth. These findings are in line with current evidence and recommendations, as we know that FV provide important micronutrients (e.g., vitamin A) for growth and development in children. That being said, contrary to our expectation and findings for vegetables/leaves and HAZ, the same study found that frequent consumption of vegetables/leaves had an inverse relationship with linear growth. The study concludes that this relationship may be due to the feeding norms of the included population of rural, breastfed Senegalese infants. In this setting, mothers have been educated to provide better feeding – including more breastmilk and diverse diets – to ill or poorly growing children. Thus, if the ill or malnourished children were provided with more breastmilk and better diets, it's possible that any positive effects on growth and nutritional status have been obscured.[45]

It is uncertain whether varying frequencies of consumption of vegetables reduces the likelihood of having low ferritin values (<20µg/L) because the certainty of this evidence is very low.[43] This study found an association between greater consumption of vegetables and greater likelihood of having low ferritin values (<20µg/L).[43] However, these findings were inconsistent as this relationship was not seen for those eating vegetables at an even higher frequency. Additionally, there was no significant relationship observed between the feeding frequency of vegetables and the likelihood of having very low iron stores (ferritin <15µg/L). Given these inconsistencies and the low quality of the data, the results should be taken with caution. Nonetheless, it is possible that these results were observed given that vegetables provide non-heme iron, which is less easily absorbed by the body unless combined with vitamin C.[56] If the

infants in this study were consuming a diet lacking in vitamin C, then their inability to absorb the iron found in the vegetables could explain the inverse relationship observed.

It is uncertain whether varying frequencies of consumption of dark green vegetables, fruit, or beans reduces the prevalence of anemia because the certainty of this evidence is very low.[46] This study reported no association between consumption of dark green vegetables, fruit (consumption in past 24 hours vs not) or beans and anemia prevalence.[46] However, when a different frequency metric was used for fruits specifically (daily vs < daily), less fruit consumption was associated with higher anemia prevalence. The lack of relationship seen for vegetables and beans can once again be explained by these foods' poor bioavailability when not consumed in combination with foods high in vitamin C, despite their iron content. In contrast, the observed relationship between fruit consumption and anemia prevalence is likely related to fruits' high content of vitamin C, or ascorbic acid, allowing for iron to be properly absorbed by the body,[57] causing a lower likelihood of anemia prevalence.[58] Furthermore, it is important to note the difference in association observed for fruit consumption and anemia prevalence when different measurement tools were used. Currently, the gold standard for undertaking quantitative dietary assessments is a weighed food record. Both FFQs and 24-hr recalls have major limitations, in particular recall bias as the participant needs to rely on their memory of past events. In this case, it may have been more appropriate to use the FFQ (daily vs < daily metric), as this provides a day-to-day account of typical food consumption, compared to a brief snapshot of foods consumed over the previous day alone. When using a 24-hr recall, it is recommended to be used at multiple time points, however this study only collected data through a 24-hr recall at one time point. A single 24-hr recall fails to capture irregular consumption of foods and can therefore underestimate the intake of some nutrients. It is likely that the use of these different tools contributed to a difference in reported associations for fruit consumption and anemia prevalence.

It is uncertain whether varying frequencies of consumption of FV improves consumption of FV later in life because the certainty of this evidence is very low.[47] This study examined higher versus lower frequency of FV consumption and changes in FV consumption later in life.[47] At both follow-up points, 36 months and 7 years of age, boys and girls showed a downward trend in vegetable consumption from 18 months of age. There was an upward trend in consumption of fruit for boys and girls between 18 and 36 months. However, at 7 years old there was no consistent trend for fruit consumption, where those who consumed lower frequencies of fruit increased their frequency of consumption, those consuming medium frequencies remained stable, and those consuming higher frequencies decreased. Overall, vegetable and fruit consumption at 18 months was positively associated with overall vegetable and fruit consumption at both 36 months and 7 years of age. These findings are consistent with the literature, where multiple studies have found that stability and tracking changes in eating habits persist beyond infancy,[59-63] suggesting the importance of setting healthy habits (e.g., frequent consumption of FV) during infancy.

In our six included studies, it was not explicitly mentioned whether study participants were food secure. However, based on sociodemographic characteristics, such as location, we believe

at least two of the included studies contained food insecure populations, if not more. This is an important point of distinction as food insecurity would play a major role in these reported health and dietary outcomes, though we have no means to assess food security status within the study populations.

There was no evidence to assess the association of differing varieties of FV or amounts of NPS with any of our primary outcomes. Furthermore, there was no evidence to assess the association between differing frequencies and varieties of FV or differing frequencies and amounts of NPS with any of our secondary outcomes: overweight/obesity, food/taste preferences later in life, markers of inflammation, markers of gut health, or adverse effects as reported by authors. Additionally, there was no evidence to assess our secondary objectives of this review, to determine differences between processed/commercial versus fresh/home-prepared fruits and vegetables regarding dietary and health outcomes.

Overall Completeness and Applicability of Evidence

In this review, we sought to determine the effects of differing frequencies, varieties, and amounts of FV and NPS consumed during the complementary feeding period on dietary and health outcomes. Our secondary objective was to review the effects of processed/commercial foods versus fresh/home prepared foods on dietary and health outcomes. Overall, we found limited and very low certainty evidence to examine the association between differing frequencies of FV intake with nutritional status, anemia prevalence, and changes/stability in FV intake later in life, and data could not establish causality. Similarly, we found limited and very low certainty evidence to examine the association between differing frequencies of NPS intake with nutritional status and anemia prevalence. We found no evidence that addressed our question of the link between processed/commercial foods versus fresh/home prepared foods and dietary and health outcomes.

Most studies were excluded on the basis of their consumption data not being reported in terms of frequency, variety or amount (n=259). However, it should be noted that there were a substantial number of studies that did report frequency, variety, or amount data for FV and NPS consumption, but the data were not linked to any health or dietary outcomes (n=24). This left us with only six included studies. Given the heterogeneity of these six studies – differing outcomes measured and metrics of frequency – the results were highly inconsistent, and we were unable to meta-analyze any of the findings. None of our included studies provided high certainty evidence in support of greater frequencies, varieties, and/or amounts of FV and NPS on dietary and health outcomes, owing partly to the fact that the vast majority were observational in nature.

Certainty of the Evidence

Overall, there is a lack of relevant and good certainty data to answer our research question. There were only six studies that met our eligibility criteria for this review, and all were

observational by design, thereby increasing the likelihood of bias. In addition to being observational in nature and the limitations that arise with this type of study design (e.g., inability to make causal conclusions), there were many methodological problems, or lack of detail provided in the methods, that led to poor certainty data. External generalizability of our findings was low to moderate based on the lack of included studies overall (n=6), but the fact that they covered most world regions, including South-East Asia, Europe, Africa, the Americas, and Western Pacific. Study sample sizes by region ranged from 205 up to 13,107 children.

Four of the studies were missing important details in their methods on factors like study population characteristics, sample size justifications, definitions and explanations on the exposure assessed (e.g., frequencies defined), and how many children were lost to follow-up. Only two studies included adjusted models in their analyses, while the remaining four did not adjust for any potential confounding variables in the relationship. This is particularly problematic for our research question as there are many potential confounding variables that could significantly affect the relationship between FV and/or NPS consumption with health and dietary outcomes. Some potential confounding variables that would have been important to control for include socioeconomic status, maternal education, food security, household environments (e.g., access to clean water and sanitation), breastfeeding status, access to health care, and childhood illness. Furthermore, there was consistently a lack of description of the types of foods consumed (e.g., use of the term vegetables and fruits rather than more specific food items) or lack of reasoning behind grouping of food. For example, one study grouped green leafy vegetables together with orange-coloured vegetables,[42] each of which provide a different set of micronutrients. Another study described frequency of FV consumption using the terms rare and frequent, but no definition was provided for what was considered rare or frequent consumption.[45]

Our rating of the overall certainty of the evidence as very low, as indicated by our GRADE Evidence Profiles, could potentially be improved upon if more studies of quasi-experimental design were conducted to answer this research question. In the instance where observational studies are most appropriate to answer the research question, it is essential that the methods employed are rigorous, including the consideration of confounding variables.

Limitations and Potential Biases in the Review Process

A key limitation in this review process was the very small sample of studies we found to include in our review (n=6). A second limitation was the lack of standardized metrics for frequency, as each author defined frequency in their own way. Given the small sample size, and lack of consistent outcomes measured, and frequency metrics used, we were unable to meta-analyze the findings, resulting in minimal to no evidence to answer our research questions.

Another limitation arose from the use of different measurement tools to determine consumption patterns of FV and/or NPS. Currently, the gold standard for such assessments is the use of a weighed food record, however these tools are labour intensive and costly, requiring many more resources. In our included studies, 24-hr recalls and FFQs were the

methods used to collect information on infants FV and NPS consumption. These tools have significant limitations, in particular they increase the risk of recall bias. The use of these tools could have led to mixed results and biased findings.

Of note, while importing references from the database CINAHL into Endnote, six of the identified references would not import. A number of tactics were attempted in order to recover these six studies (e.g., splitting the set of references up into smaller chunks for importation, etc.) however we were unsuccessful in doing so. Given that we searched a large number of databases, including CINAHL, it is likely that these six studies were captured by another database. Nonetheless, it's possible that these studies were missed and could have biased our findings. We were also unable to retrieve full texts from 12 studies that may have qualified for inclusion into this review.

Given our search strategy was comprehensive (it consisted of a complex mixture of keywords and MeSH terms, and it was run in multiple databases with no language restrictions, and grey literature was searched) it is very unlikely that location or language bias occurred in our review. Furthermore, a detailed and defined selection criterion was outlined and decided prior to beginning this review, thus selection bias is also unlikely to have occurred.

Although publication bias is hard to avoid, several of our included studies reported inconsistent findings with the literature or no association (no statistical relevance) was found between exposure and outcome. For this reason, it seems relatively unlikely that there is major risk of publication bias in this review.

Agreements and Disagreements with Other Studies or Reviews

To our knowledge, this is the first study to systematically review the association between differing frequencies, varieties, and/or amounts of FV or NPS consumption with dietary and health outcomes.

Our review found evidence for an association between less frequent consumption of FV and greater prevalence of stunting and underweight.[44] Similarly, we found evidence for an association between less frequent consumption of FV and lower HAZ score, and more frequent consumption of fruit with increased linear growth.[45] We also found evidence for an association between less fruit consumption and greater prevalence of anemia.[46] We could not find any other studies that have examined any of these relationships, however in all four cases these finding were in line with our expectations, as we know that fruits and vegetables provide important sources of nutrients in the diet (e.g., potassium, folate, fiber, vitamin A, vitamin C, vitamin K, etc.), thus if consumed less frequently could be associated with higher prevalence of poor dietary and health outcomes (e.g., stunting, underweight, anemia, and HAZ). The inverse situation is also true, in which more frequently consumed FV could be associated with improved dietary and health outcomes (e.g., linear growth). Furthermore, we found evidence that overall FV consumption at 18 months of age was positively associated with overall FV

consumption later in life (36 months and 7 years of age).[47] This finding is consistent with other studies that have found that eating habits early in infancy persist well into later years of life, highlighting the importance of setting healthy habits at an early age.[59-63]

Four findings from our review were unexpected with regards to our understanding of these topics. Firstly, Ahmad 2018[42] found no association between varying frequencies of consumption of FV or NPS on the prevalence of wasting, stunting, and underweight. Secondly, Ntab 2005[45] found an inverse relationship between more frequent consumption of vegetables and linear growth. We were unable to find any previous studies that have examined this relationship, however, the study concludes that this finding may be due to the feeding norms of the included population, where caregivers have been told to feed more breastmilk and diverse diets when their child is ill. In this case, it's possible that any positive effects of growth and nutritional status in the study participants were obscured by the child's ill health status.

Wandel 1996[43] demonstrated inconsistent findings when examining the association between vegetable consumption and ferritin concentration. An inverse relationship was observed, where greater consumption of vegetables was associated with a greater likelihood of having a low ferritin concentration (<20µg/L). However, this relationship did not hold when looking at even greater frequencies of vegetable consumption, nor when looking at the likelihood of having very low iron stores (<15µg/L). Similarly, Silva 2007[46] found no association between consumption of FV and NPS with the prevalence of anemia. In both cases, it is possible that these results were impacted by the fact that vegetables and NPS provide non-heme iron, which is less easily absorbed by the body unless combined with vitamin C,[56] despite their iron content.

AUTHOR'S CONCLUSIONS

Implications for Practice and Policy

The aim of this review was to determine the optimal frequency, variety, or amounts of FV and/or NPS consumption for beneficial dietary and health outcomes. The findings were planned to inform the development of the updated WHO IYCF guidelines, in addition to future potential policies and programmes relating to complementary feeding. However, overall we have found a lack of relevant and good quality data for there to be implications for practice and policy at this time. Instead, this review has highlighted a major gap in the evidence, and more research is required.

Implications for Research

This review highlights the need for much more research to be conducted that examines various frequencies, varieties, and amounts of FV and NPS consumption during the complementary feeding period and their association with dietary and health outcomes. Ideally, such research

would take the shape of good quality experimental and quasi-experimental studies (e.g., controlled before-after studies), or longitudinal cohort studies whereby the internal comparator groups could be used to properly assess outcomes across groups. Furthermore, it is imperative that potential confounders are addressed in the study analyses.

Given the nutrition transition that has now reached a global scale, it is also crucial that future research addresses our secondary objective of determining the health or nutrition effects of processed/commercial versus fresh/home-prepared foods in the context of optimal frequency, variety, or amounts of complementary foods, including FV and NPS).

REFERENCES

1. PAHO, WHO. Guiding principles for complementary feeding of the breastfed child. Washington, DC: Pan American Health Organization. Published 2003.

2. Victora CG, Adair L, Fall C, et al. Maternal and child undernutrition: Consequences for adult health and human capital. *Lancet.* 2008; 371(9609):340-357.

3. UNICEF. The State of the World's Children 2019. Children, food and nutrition: Growing well in a changing world. UNICEF, New York. Published 2019.

4. Hoddinott J, Behrman JR, Maluccio JA, et al. Adult consequences of growth failure in early childhood. *Am J Clin Nutr.* 2013; 98(5):1170-1178.

5. Martorell R. Improved nutrition in the first 1000 days and adult human capital and health. *Am J Hum Biol.* 2017; 29(2).

6. World Health Organization. Malnutrition: key facts. Published 2020. Retrieved from: https://www.who.int/news-room/fact-sheets/detail/malnutrition

7. Eaton JC, Rothpletz-Puglia P, Dreker MR, Iannotti L, Lutter C, Kaganda J, et al. Effectiveness of animal-source foods for supporting optimal growth and development in children 6 to 59 months of age. *Cochrane Database Syst. Rev.* 2019; 2(2): CD012818.

8. Black RE, Allen LH, Bhutta ZA, Caulfield LE, Onis M, Ezzati M, et al. Maternal and child undernutrition: global and regional exposures and health consequences. *Lancet*. 2008; 371(9608):243-60.

9. WHO. Complementary Feeding: Family foods for breastfed children. Published 2000. Retrieved from: https://apps.who.int/iris/bitstream/handle/10665/66389/WHO_NHD_00.1.pdf?sequence=1

10. Issaka AI, Agho KE, Page AN, Burns PL, Stevens GJ, & Dibley MJ. The problem of suboptimal complementary feeding practices in West Africa: what is the way forward? *Matern Child Nutr.* 2015; Suppl 1 (Suppl 1): 53-60.

11. Lartey A. Maternal and child nutrition in Sub-Saharan Africa: challenges and interventions. *Proceedings of the Nutrition Society.* 2008; 67: 105– 108.

12. Joshi N, Agho KE, Dibley MJ, Senarath U, & Tiwari K. Determinants of inappropriate complementary feeding practices in young children in Nepal: secondary data analysis of Demographic and Health Survey 2006. *Maternal & Child Nutrition*. 2012; 8: 45– 59.

13. Kabir I, Khanam M, Agho KE, Mihrshahi S, Dibley MJ, & Roy SK. Determinants of inappropriate complementary feeding practices in infant and young children in Bangladesh: secondary data analysis of Demographic Health Survey 2007. *Maternal & Child Nutrition*. 2012; 8:11– 27.

14. Menon P. The crisis of poor complementary feeding in South Asia: where next? *Maternal & Child Nutrition*. 2012; 8: 1– 4.

15. WHO. Guiding principles for feeding non-breastfed children 6-24 months of age. Published 2005. Retrieved from: https://apps.who.int/iris/bitstream/handle/10665/43281/9241593431.pdf?sequence=1.

16. WHO. Indicators for assessing infant and young child feeding practices. Geneva. Published 2008. Retrieved from: https://apps.who.int/iris/bitstream/handle/10665/43895/9789241596664_eng.pdf;jsessionid=2EA7A0B40867BEE1E27D839F1F0DEC28?sequence=1.

17. Lutter CK, Grummer-Strawn L, & Rogers L. Complementary feeding of infants and young children 6 to 23 months of age. *Nutrition Reviews*. 2021; nuaa143.

18. WHO. Infant and Young Child Feeding: Model Chapter for Textbooks for Medical Students and Allied Health Professionals. Published 2009. Retrieved from: https://www.ncbi.nlm.nih.gov/books/NBK148957/

19. Allen LH, Gillespie S. What works? A review of the efficacy and effectiveness of nutrition interventions. ACC/SCN Nutrition Policy Paper No. 19. ACC/SCN: Geneva in collaboration with the Asian Development Bank, Manila, 2001.

20. Rivera J, Santizo MC, & Hurtado E. Diseño y evaluación de un programa educativo para mejorar las prácticas de alimentación en niños de 6 a 24 meses de edad en comunidades rurales de Guatemala. (Design and evaluation of an educational programme to improve feeding practices of children 6 to 24 months of age in rural communities of Guatemala). Organización Panamericana de la Salud, 1998.

21. Nebeling L. Phytochemicals: the color of a healthy diet. Pediatric Basics. *J Ped Nutr Dev*. 2002; 98:2–9.

22. McGuire S. U.S. Department of Agriculture and U.S. Department of Health and Human Services, Dietary Guidelines for Americans, 2010. 7th Edition, Washington, DC: U.S. Government Printing Office, January 2011. *Advances in Nutrition*. 2011; 2(3):293-294.

23. Grimm KA, Kim SA, Yaroch AL, et al. Fruit and vegetable intake during infancy and early childhood. *Pediatrics.* 2014; 134: S63–9.

24. Birch LL, Gunder L, Grimm-Thomas K, & Laing DG. Infants' consumption of a new food enhances acceptance of similar foods. *Appetite.* 1998;30(3):283-295.

25. Sullivan SA, & Birch LL. Infant dietary experience and acceptance of solid foods. *Pediatrics.* 1994;93(2):271-277.

26. Hetherington MM, Schwartz C, Madrelle J, et al. A step-by-step introduction to vegetables at the beginning of complementary feeding. The effects of early and repeated exposure. *Appetite.* 2015; 84:280-290.

27. Coulthard H, Harris G, Emmett P. Delayed introduction of lumpy foods to children during the complementary feeding period affects child's food acceptance and feeding at 7 years of age. *Matern Child Nutr.* 2009;5(1):75-85.

28. Chambers L. Complementary feeding: Vegetables first, frequently and in variety. *Nutrition Bulletin*. 2016; 41: 142-146.

29. Coulthard H, Harris G, & Emmett P. Long-term consequences of early fruit and vegetable feeding practices in the United Kingdom. *Public Health Nutrition.* 2010;13: 2044–51.

30. Amezdroz E, Carpenter L, O'Callaghan E, Johnson S, & Waters E. Transition from milks to the introduction of solid foods across the first 2 years of life: Findings from an australian birth cohort study. *J Hum Nutr Diet.* 2015;28(4):375-383.

31. Deming DM, Afeiche MC, Reidy KC, Eldridge AL, & Villalpando-Carrión S. Early feeding patterns among mexican babies: Findings from the 2012 national health and nutrition survey and implications for health and obesity prevention. *BMC Nutr.* 2012; 1:40.

32. G. B. D. Diet Collaborators. Health effects of dietary risks in 195 countries, 1990-2017: A systematic analysis for the global burden of disease study 2017. *Lancet.* 2019*;* 393(10184):1958-1972.

33. Stevens GA, Finucane MM, De-Regil LM, Paciorek CJ, Flaxman SR, Branca F, et al. Global, regional, and national trends in haemoglobin concentration and prevalence of total and severe anaemia in children and pregnant and non-pregnant women for 1995-2011: a systematic analysis of population-representative data. *Lancet*. 2013; 1(1): E16-E25.

34. Alberta Health Services. Nutrition Guideline Healthy Infants and Young Children Introduction of Complementary Foods. Published 2015. Retrieved from:

https://www.albertahealthservices.ca/assets/info/nutrition/if-nfs-ng-healthy-infants-complementary-feeding-introduction-complementary-foods.pdf

35. Abeshu MA, Lelisa A, & Geleta B. Complementary Feeding: Review of Recommendations, Feeding Practices, and Adequacy of Homemade Complementary Food Preparations in Developing Countries - Lessons from Ethiopia. *Front Nutr*. 2016; 3:41.

36. Dewey KG, Cohen RJ, Brown KH, & Rivera LL. Age of introduction of complementary food and growth of term, low birth weight breastfed infants: a randomized intervention study in Honduras. *Am J Clin Nutr*. 1999; 69: 679-86.

37. Cohen RJ, Brown KH, Canahuati J, Rivera LL, & Dewey KG. Effects of age of introduction of complementary foods on infant breast milk intake, total energy intake, and growth: a randomized intervention study in Honduras. *Lancet*. 1994; 344:288-93.

38. WHO. WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height, and body mass index-for-age: methods and development. WHO, France. Published 2006.

39. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. 2019; 366: l4898.

40. NIH. Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies. Retrieved from: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools

41. Sterne JAC, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomized studies of interventions. *BMJ*. 2016; 355: i4919.

42. Ahmad A, Madanijah S, Dwiriani CM, & Kolopaking R. Complementary feeding practices and nutritional status of children 6-23 months old: formative study in Aceh, Indonesia. *Nutrition Research and Practice*. 2018; 12(6): 512-520.

43. Wandel M, Fagerli RA, Olsen PT, Borch-Iohnsen B, & Ek J. Iron status and weaning practices among Norwegian and immigrant infants. *Nutrition Research*. 1996; 16(2): 251-265.

44. Chang S, Chen C, He W, & Wang Y. Analysis on the changes of nutritional status among Chinese infants and young children -- the improvement of complementary feeding. *Journal of Hygiene Research.* 2007; 36(2):207-9.

45. Ntab B, Simondon KB, Milet J, Cissé B, Sokhna C, Boulanger D, et al. A Young Child Feeding Index Is Not Associated with Either Height-for-Age or Height Velocity in Rural Senegalese Children. *The Journal of Nutrition*. 2004; 135(3): 457-464.

46. Silva DG, Priore SE, & Franceschini SCC. Risk factors for anemia in infants assisted by public health services: the importance of feeding practices and iron supplementation. *J Pediatr*. 2007; 83(2): 149-156.

47. Bjelland M, Brantsaeter AL, Haugen M, Meltzer HM, Nystad W, & Andersen LF. Changes and tracking of fruit, vegetables and sugar-sweetened beverages intake from 18 months to 7 years in the Norwegian mother and child cohort study. *BMC Public Health*. 2013; 13:793.

48. WHO. Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity, 2011.

49. NCT02634749. OTIS- Optimized Complementary Feeding Study (OTIS). https://clinicaltrials.gov/ct2/show/record/NCT02634749

50. PACTR201906819960554. Efficacy of a strategy combining the promotion of responsive feeding with the daily consumption of a fortified cereal-based blend to reduce anemia, micronutrient deficiencies and stunting in infants from Amparafaravola District, Madagascar. https://pactr.samrc.ac.za/TrialDisplay.aspx?TrialID=5972

51. Teshome GB, Whiting SJ, Green TJ, Mulualem D, & Henry C. Scaled-up nutrition education on pulse-cereal complementary food practice in Ethiopia: a cluster-randomized trial. *BMC Public Health*. 2020; 20:1437.

52. Mesfin A, Henry C, Girma M, & Whiting SJ. Use of pulse crops in complementary feeding of 6-23-month-old infants and young children in Taba Kebele, Damot Gale District, Southern Ethiopia. *Journal of Public Health in Africa.* 2015; 6(1):357.

53. Kimiywe J, & Chege P. Complementary feeding practices and nutritional status of children 6-23 months in Kitui County, Kenya. *J Appl Biosci*. 2015; 85:7881-90.

54. Korir JK. Determinants of complementary feeding practices and nutritional status of children 6-23 months old in Korogocho slum Nairobi County, Kenya [master's thesis]. *Nairobi: Kenyatta University*. Published 2013.

55. Krasevec J, An X, Kumapley R, Bégin F, & Frongillo EA. Diet quality and risk of stunting among infants and young children in low- and middle-income countries. *Matern Child Nutr*. 2017; 13 Suppl 2 (Suppl 2): e12430.

56. Gillooly M, Bothwell TH, Torrance JD, MacPhail AP, Derman DP, Bezwoda WR, et al. The effects of organic acids, phytates, and polyphenols on the absorption of iron from vegetables. *Br J Nutr*. 1983; 49(3): 331-42.

57. Fairweather-Tait S, Fox T, Wharf SG, & Eagles J. The bioavailability of iron in different weaning foods and the enhancing effect of a fruit drink containing ascorbic acid. *Pediatr Res*. 1995; 37(4 Pt 1):389-94.

58. DeMaeyer EM, Dallman P, Gurney JM, Hallberg L, Sood SK, et al. Preventing and controlling iron deficiency anaemia through primary health care: a guide for health administrators and programme managers. World Health Organization. Published 1989. Retrieved from: https://apps.who.int/iris/handle/10665/39849

59. Madruga SW, Araujo CLP, Bertoldi AD, & Neutzling MB. Tracking of dietary patterns from childhood to adolescence. *Rev Saude Publica*. 2012; 46:376–386.

60. Stein AD, Shea S, Basch CE, Contento IR, & Zybert P. Variability and tracking of nutrient intakes of preschool children based on multiple administrations of the 24-hour dietary recall. *Am J Epidemiol*. 1991; 134:1427–1437.

61. Boulton TJ, Magarey AM, & Cockington RA. Tracking of serum lipids and dietary energy, fat and calcium intake from 1 to 15 years. *Acta Paediatr*. 1995; 84:1050–1055.

62. Singer MR, Moore LL, Garrahie EJ, & Ellison RC. The tracking of nutrient intake in young children: the Framingham Children's study. *Am J Public Health*. 1995; 85:1673–1677.

63. Robinson S, Marriott L, Poole J, Crozier S, Borland S, Lawrence W, et al. Southampton Women's Survey Study Group: Dietary patterns in infancy: the importance of maternal and family influences on feeding practice. *Br J Nutr*. 2007; 98:1029–1037.

APPENDICES

Appendix A- Medline Search Strategy

Search Number	**Search Terms**
Search Set 1	**Child Terms**
1	exp Infant, Newborn/
2	exp Infant/
3	child/ or child, preschool/
4	(infan* or babies or baby or newborn? or neonat* or toddler? or child*). tw,kf.
5	**or/1-4**
Search Set 2	**Complementary Feeding Terms**
6	Weaning/
7	Infant Nutritional Physiological Phenomena/
8	Feeding Behavior/
9	Eating/
10	complementary feed*. tw,kf.
11	(feed* or food? or diet* or eat* or wean* or consume? or consuming or consumption). tw,kf.
12	**or/6-11**
Search Set 3	**Food Terms**
13	Fruit/
14	Vegetables/
15	meat/ or meat products/ or exp meat proteins/ or exp poultry/ or exp red meat/ or exp seafood/
16	dairy products/ or exp cultured milk products/ or exp milk/
17	nuts/ or seeds/
18	Fabaceae/ or Phaseolus/

19	**or/13-18**
20	(frequen* or quantit* or regular* or amount* or variet* or divers* or type*).tw,kf.
21	**19 and 20**
22	((leafy or vitamin-A rich or vitamin A rich or nutrient dense or nutrient-dense or nutrient rich or nutrient-rich) adj5 (frequen* or quantit* or amount* or variet* or divers* or type*)).tw,kf.
23	((fruit? or vegetable? or greens or broccoli or apricot* or papaya* or avocado* or cabbage* or squash* or banana* or orange* or carrot* or mango*) adj5 (frequen* or quantit* or amount* or variet* or divers* or type*)).tw,kf.
24	((meat* or beef or pork or lamb or goat* or game or poultry or chicken* or camel? or venison or egg* or insect* or caterpillar* or spider* or beetle* or termite* or ant or ants or animal-source? or flesh-food? or flesh food?) adj5 (frequen* or quantit* or amount* or variet* or divers* or type*)).tw,kf.
25	((seafood or sea-food or shellfish* or fish* or marine*) adj5 (frequen* or quantit* or amount* or variet* or divers* or type*)).tw,kf.
26	((dairy or milk or cheese or yogurt or butter or kefir or kephir or bulgaros) adj5 (frequen* or quantit* or amount* or variet* or divers* or type*)).tw,kf.
27	((legume? or pulse? or nut? or seed? or bean? or lentil? or chickpea? or pea? or cowpea? or soybean* or chestnut*) adj5 (frequen* or quantit* or amount* or variet* or divers* or type*)).tw,kf.
28	**or/21-27**
29	**5 and 12 and 28**

Appendix B- NIH, ROB-2, and ROBINS-I Tools for Quality Assessments

A. NIH Tool for Observational Cohort and Cross-Sectional Studies

Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies —

Criteria	Yes	No	Other (CD, NR, NA)*
1. Was the research question or objective in this paper clearly stated?			
2. Was the study population clearly specified and defined?			
3. Was the participation rate of eligible persons at least 50%?			
4. Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion criteria for being in the study prespecified and applied uniformly to all participants?			
5. Was a sample size justification, power description, or variance and effect estimates provided?			
6. For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured?			
7. Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed?			
8. For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (e.g., categories of exposure, or exposure measured as continuous variable)?			
9. Were the exposure measures (independent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?			
10. Was the exposure(s) assessed more than once over time?			
11. Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?			
12. Were the outcome assessors blinded to the exposure status of participants?			
13. Was loss to follow-up after baseline 20% or less?			
14. Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcome(s)?			

B. ROB-2 Tool for Randomized-Controlled Trials

Domain 1: Risk of bias arising from the randomization process

Signalling questions	Elaboration	Response options
1.1 Was the allocation sequence random?	Answer 'Yes' if a random component was used in the sequence generation process. Examples include computer-generated random numbers; reference to a random number table; coin tossing; shuffling cards or envelopes; throwing dice; or drawing lots. Minimization is generally implemented with a random element (at least when the scores are equal), so an allocation sequence that is generated using minimization should generally be considered to be random. Answer 'No' if no random element was used in generating the allocation sequence or the sequence is predictable. Examples include alternation; methods based on dates (of birth or admission); patient record numbers; allocation decisions made by clinicians or participants; allocation based on the availability of the intervention; or any other systematic or haphazard method. Answer 'No information' if the only information about randomization methods is a statement that the study is randomized. In some situations a judgement may be made to answer 'Probably no' or 'Probably yes'. For example, , in the context of a large trial run by an experienced clinical trials unit, absence of specific information about generation of the randomization sequence, in a paper published in a journal with rigorously enforced word count limits, is likely to result in a response of 'Probably yes' rather than 'No information'. Alternatively, if other (contemporary) trials by the same investigator team have clearly used non-random sequences, it might be reasonable to assume that the current study was done using similar methods.	Y/PY/PN/N/NI
1.2 Was the allocation sequence concealed until participants were enrolled and assigned to interventions?	Answer 'Yes' if the trial used any form of remote or centrally administered method to allocate interventions to participants, where the process of allocation is controlled by an external unit or organization, independent of the enrolment personnel (e.g. independent central pharmacy, telephone or internet-based randomization service providers). Answer 'Yes' if envelopes or drug containers were used appropriately. Envelopes should be opaque, sequentially numbered, sealed with a tamper-proof seal and opened only after the envelope has been irreversibly assigned to the participant. Drug containers should be sequentially numbered and of identical appearance, and dispensed or administered only after they have been irreversibly assigned to the participant. This level of detail is rarely provided in reports, and a judgement may be required to justify an answer of 'Probably yes' or 'Probably no'. Answer 'No' if there is reason to suspect that the enrolling investigator or the participant had knowledge of the forthcoming allocation.	Y/PY/PN/N/NI
1.3 Did baseline differences between intervention groups suggest a problem with the randomization process?	*Note that differences that are compatible with chance do not lead to a risk of bias. A small number of differences identified as 'statistically significant' at the conventional 0.05 threshold should usually be considered to be compatible with chance.* Answer 'No' if no imbalances are apparent or if any observed imbalances are compatible with chance. Answer 'Yes' if there are imbalances that indicate problems with the randomization process, including: (1) substantial differences between intervention group sizes, compared with the intended allocation ratio; or (2) a substantial excess in statistically significant differences in baseline characteristics between intervention groups, beyond that expected by chance; or (3) imbalance in one or more key prognostic factors, or baseline measures of outcome variables, that is very unlikely to be due to chance and for which the between-group difference is big enough to result in bias in the intervention effect estimate. Also answer 'Yes' if there are other reasons to suspect that the randomization process was problematic: (4) excessive similarity in baseline characteristics that is not compatible with chance. Answer 'No information' when there is no *useful* baseline information available (e.g. abstracts, or studies that reported only baseline characteristics of participants in the final analysis). The answer to this question should not influence answers to questions 1.1 or 1.2. For example, if the trial has large baseline imbalances, but authors report adequate randomization methods, questions 1.1 and 1.2 should still be answered on the basis of the reported adequate methods, and any concerns about the imbalance should be raised in the answer to the question 1.3 and reflected in the domain-level risk-of-bias judgement. Trialists may undertake analyses that attempt to deal with flawed randomization by controlling for imbalances in prognostic factors at baseline. To remove the risk of bias caused by problems in the randomization process, it would be necessary to know, and measure, all the prognostic factors that were imbalanced at baseline. It is unlikely that all important prognostic factors are known and measured, so such analyses will at best reduce the risk of bias. If review authors wish to assess the risk of bias in a trial that controlled for baseline imbalances in order to mitigate failures of randomization, the study should be assessed using the ROBINS-I tool.	Y/PY/PN/N/NI
Risk-of-bias judgement	See algorithm.	Low / High / Some concerns

Domain 2: Risk of bias due to deviations from the intended interventions (*effect of assignment to intervention*)

Signalling questions	Elaboration	Response options
2.1. Were participants aware of their assigned intervention during the trial?	If participants are aware of their assigned intervention it is more likely that health-related behaviours will differ between the intervention groups. Blinding participants, most commonly through use of a placebo or sham intervention, may prevent such differences. If participants experienced side effects or toxicities that they knew to be specific to one of the interventions, answer this question 'Yes' or 'Probably yes'.	Y/PY/PN/N/NI
2.2. Were carers and people delivering the interventions aware of participants' assigned intervention during the trial?	If carers or people delivering the interventions are aware of the assigned intervention then its implementation, or administration of non-protocol interventions, may differ between the intervention groups. Blinding may prevent such differences. If participants experienced side effects or toxicities that carers or people delivering the interventions knew to be specific to one of the interventions, answer question 'Yes' or 'Probably yes'. If randomized allocation was not concealed, then it is likely that carers and people delivering the interventions were aware of participants' assigned intervention during the trial.	Y/PY/PN/N/NI
2.3. If Y/PY/NI to 2.1 or 2.2: Were there deviations from the intended intervention that arose because of the trial context?	For the effect of assignment to intervention, this domain assesses problems that arise when changes from assigned intervention that are inconsistent with the trial protocol arose because of the trial context. We use the term **trial context** to refer to effects of recruitment and engagement activities on trial participants and when trial personnel (carers or people delivering the interventions) undermine the implementation of the trial protocol in ways that would not happen outside the trial. For example, the process of securing informed consent may lead participants subsequently assigned to the comparator group to feel unlucky and therefore seek the experimental intervention, or other interventions that improve their prognosis. Answer 'Yes' or 'Probably yes' **only** if there is evidence, or strong reason to believe, that the trial context led to failure to implement the protocol interventions or to implementation of interventions not allowed by the protocol. Answer 'No' or 'Probably no' if there were changes from assigned intervention that are inconsistent with the trial protocol, such as non-adherence to intervention, but these are consistent with what could occur outside the trial context. Answer 'No' or 'Probably no' for changes to intervention that are consistent with the trial protocol, for example cessation of a drug intervention because of acute toxicity or use of additional interventions whose aim is to treat consequences of one of the intended interventions. If blinding is compromised because participants report side effects or toxicities that are specific to one of the interventions, answer 'Yes' or 'Probably yes' only if there were changes from assigned intervention that are inconsistent with the trial protocol and arose because of the trial context. The answer 'No information' may be appropriate, because trialists do not always report whether deviations arose because of the trial context.	NA/Y/PY/PN/N/NI
2.4 If Y/PY to 2.3: Were these deviations likely to have affected the outcome?	Changes from assigned intervention that are inconsistent with the trial protocol and arose because of the trial context will impact on the intervention effect estimate if they affect the outcome, but not otherwise.	NA/Y/PY/PN/N/NI

2.5. If Y/PY/NI to 2.4: Were these deviations from intended intervention balanced between groups?	Changes from assigned intervention that are inconsistent with the trial protocol and arose because of the trial context are more likely to impact on the intervention effect estimate if they are not balanced between the intervention groups.	NA/Y/PY/PN/N/NI
2.6 Was an appropriate analysis used to estimate the effect of assignment to intervention?	Both intention-to-treat (ITT) analyses and modified intention-to-treat (mITT) analyses excluding participants with missing outcome data should be considered appropriate. Both naïve 'per-protocol' analyses (excluding trial participants who did not receive their assigned intervention) and 'as treated' analyses (in which trial participants are grouped according to the intervention that they received, rather than according to their assigned intervention) should be considered inappropriate. Analyses excluding eligible trial participants post-randomization should also be considered inappropriate, but post-randomization exclusions of ineligible participants (when eligibility was not confirmed until after randomization, and could not have been influenced by intervention group assignment) can be considered appropriate.	Y/PY/PN/N/NI
2.7 If N/PN/NI to 2.6: Was there potential for a substantial impact (on the result) of the failure to analyse participants in the group to which they were randomized?	This question addresses whether the number of participants who were analysed in the wrong intervention group, or excluded from the analysis, was sufficient that there could have been a substantial impact on the result. It is not possible to specify a precise rule: there may be potential for substantial impact even if fewer than 5% of participants were analysed in the wrong group or excluded, if the outcome is rare or if exclusions are strongly related to prognostic factors.	NA/Y/PY/PN/N/NI
Risk-of-bias judgement	See algorithm.	Low / High / Some concerns
Optional: What is the predicted direction of bias due to deviations from intended interventions?	If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions.	NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Domain 3: Risk of bias due to missing outcome data

Signalling questions	Elaboration	Response options
3.1 Were data for this outcome available for all, or nearly all, participants randomized?	The appropriate study population for an analysis of the intention to treat effect is all randomized participants. "Nearly all" should be interpreted as that the number of participants with missing outcome data is sufficiently small that their outcomes, whatever they were, could have made no important difference to the estimated effect of intervention. For continuous outcomes, availability of data from 95% of the participants will often be sufficient. For dichotomous outcomes, the proportion required is directly linked to the risk of the event. If the observed number of events is much greater than the number of participants with missing outcome data, the bias would necessarily be small. Only answer 'No information' if the trial report provides no information about the extent of missing outcome data. This situation will usually lead to a judgement that there is a high risk of bias due to missing outcome data. Note that imputed data should be regarded as missing data, and not considered as 'outcome data' in the context of this question.	Y/PY/PN/N/NI
3.2 If N/PN/NI to 3.1: Is there evidence that the result was not biased by missing outcome data?	Evidence that the result was not biased by missing outcome data may come from: (1) analysis methods that correct for bias; or (2) sensitivity analyses showing that results are little changed under a range of plausible assumptions about the relationship between missingness in the outcome and its true value. However, imputing the outcome variable, either through methods such as 'last-observation-carried-forward' or via multiple imputation based only on intervention group, should not be assumed to correct for bias due to missing outcome data.	NA/Y/PY/PN/N
3.3 If N/PN to 3.2: Could missingness in the outcome depend on its true value?	If loss to follow up, or withdrawal from the study, could be related to participants' health status, then it is possible that missingness in the outcome was influenced by its true value. However, if all missing outcome data occurred for documented reasons that are unrelated to the outcome then the risk of bias due to missing outcome data will be low (for example, failure of a measuring device or interruptions to routine data collection). In time-to-event analyses, participants censored during trial follow-up, for example because they withdrew from the study, should be regarded as having missing outcome data, even though some of their follow up is included in the analysis. Note that such participants may be shown as included in analyses in CONSORT flow diagrams.	NA/Y/PY/PN/N/NI

3.4 If Y/PY/NI to 3.3: Is it likely that missingness in the outcome depended on its true value?	This question distinguishes between situations in which (i) missingness in the outcome could depend on its true value (assessed as 'Some concerns') from those in which (ii) it is likely that missingness in the outcome depended on its true value (assessed as 'High risk of bias'). Five reasons for answering 'Yes' are: 1. Differences between intervention groups in the proportions of missing outcome data. If there is a difference between the effects of the experimental and comparator interventions on the outcome, and the missingness in the outcome is influenced by its true value, then the proportions of missing outcome data are likely to differ between intervention groups. Such a difference suggests a risk of bias due to missing outcome data, because the trial result will be sensitive to missingness in the outcome being related to its true value. For time-to-event-data, the analogue is that rates of censoring (loss to follow-up) differ between the intervention groups. 2. Reported reasons for missing outcome data provide evidence that missingness in the outcome depends on its true value; 3. Reported reasons for missing outcome data differ between the intervention groups; 4. The circumstances of the trial make it likely that missingness in the outcome depends on its true value. For example, in trials of interventions to treat schizophrenia it is widely understood that continuing symptoms make drop out more likely. 5. In time-to-event analyses, participants' follow up is censored when they stop or change their assigned intervention, for example because of drug toxicity or, in cancer trials, when participants switch to second-line chemotherapy. Answer 'No' if the analysis accounted for participant characteristics that are likely to explain the relationship between missingness in the outcome and its true value.	NA/Y/PY/PN/N/NI
Risk-of-bias judgement	See algorithm.	Low / High / Some concerns
Optional: What is the predicted direction of bias due to missing outcome data?	If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions.	NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Domain 4: Risk of bias in measurement of the outcome

Signalling questions	Elaboration	Response options
4.1 Was the method of measuring the outcome inappropriate?	This question aims to identify methods of outcome measurement (data collection) that are unsuitable for the outcome they are intended to evaluate. The question *does not* aim to assess whether the choice of outcome being evaluated was sensible (e.g. because it is a surrogate or proxy for the main outcome of interest). In most circumstances, for pre-specified outcomes, the answer to this question will be 'No' or 'Probably no'. Answer 'Yes' or 'Probably yes' if the method of measuring the outcome is inappropriate, for example because: (1) it is unlikely to be sensitive to plausible intervention effects (e.g. important ranges of outcome values fall outside levels that are detectable using the measurement method); or (2) the measurement instrument has been demonstrated to have poor validity.	Y/PY/PN/N/NI
4.2 Could measurement or ascertainment of the outcome have differed between intervention groups?	Comparable methods of outcome measurement (data collection) involve the same measurement methods and thresholds, used at comparable time points. Differences between intervention groups may arise because of 'diagnostic detection bias' in the context of passive collection of outcome data, or if an intervention involves additional visits to a healthcare provider, leading to additional opportunities for outcome events to be identified.	Y/PY/PN/N/NI
4.3 If N/PN/NI to 4.1 and 4.2: Were outcome assessors aware of the intervention received by study participants?	Answer 'No' if outcome assessors were blinded to intervention status. For participant-reported outcomes, the outcome assessor is the study participant.	NA/Y/PY/PN/N/NI
4.4 If Y/PY/NI to 4.3: Could assessment of the outcome have been influenced by knowledge of intervention received?	Knowledge of the assigned intervention could influence participant-reported outcomes (such as level of pain), observer-reported outcomes involving some judgement, and intervention provider decision outcomes. They are unlikely to influence observer-reported outcomes that do not involve judgement, for example all-cause mortality.	NA/Y/PY/PN/N/NI

4.5 If Y/PY/NI to 4.4: Is it likely that assessment of the outcome was influenced by knowledge of intervention received?	This question distinguishes between situations in which (i) knowledge of intervention status could have influenced outcome assessment but there is no reason to believe that it did (assessed as 'Some concerns') from those in which (ii) knowledge of intervention status was likely to influence outcome assessment (assessed as 'High'). When there are strong levels of belief in either beneficial or harmful effects of the intervention, it is more likely that the outcome was influenced by knowledge of the intervention received. Examples may include patient-reported symptoms in trials of homeopathy, or assessments of recovery of function by a physiotherapist who delivered the intervention.	NA/Y/PY/PN/N/NI
Risk-of-bias judgement	See algorithm.	Low / High / Some concerns
Optional: What is the predicted direction of bias in measurement of the outcome?	If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions.	NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Domain 5: Risk of bias in selection of the reported result

Signalling questions	Elaboration	Response options
5.1 Were the data that produced this result analysed in accordance with a pre-specified analysis plan that was finalized before unblinded outcome data were available for analysis?	If the researchers' pre-specified intentions are available in sufficient detail, then planned outcome measurements and analyses can be compared with those presented in the published report(s). To avoid the possibility of selection of the reported result, finalization of the analysis intentions must precede availability of unblinded outcome data to the trial investigators. Changes to analysis plans that were made before unblinded outcome data were available, or that were clearly unrelated to the results (e.g. due to a broken machine making data collection impossible) do not raise concerns about bias in selection of the reported result.	Y/PY/PN/N/NI
Is the numerical result being assessed likely to have been selected, on the basis of the results, from...		
5.2. ... multiple eligible outcome measurements (e.g. scales, definitions, time points) within the outcome domain?	A particular outcome domain (i.e. a true state or endpoint of interest) may be measured in multiple ways. For example, the domain pain may be measured using multiple scales (e.g. a visual analogue scale and the McGill Pain Questionnaire), each at multiple time points (e.g. 3, 6 and 12 weeks post-treatment). If multiple measurements were made, but only one or a subset is reported on the basis of the results (e.g. statistical significance), there is a high risk of bias in the fully reported result. Attention should be restricted to outcome measurements that are eligible for consideration by the RoB 2 tool user. For example, if only a result using a specific measurement scale is eligible for inclusion in a meta-analysis (e.g. Hamilton Depression Rating Scale), and this is reported by the trial, then there would not be an issue of selection even if this result was reported (on the basis of the results) in preference to the result from a different measurement scale (e.g. Beck Depression Inventory). Answer 'Yes' or 'Probably yes' if: There is clear evidence (usually through examination of a trial protocol or statistical analysis plan) that a domain was measured in multiple eligible ways, but data for only one or a subset of measures is fully reported (without justification), and the fully reported result is likely to have been selected on the basis of the results. Selection on the basis of the results can arise from a desire for findings to be newsworthy, sufficiently noteworthy to merit publication, or to confirm a prior hypothesis. For example, trialists who have a preconception, or vested interest in showing, that an	Y/PY/PN/N/NI

	experimental intervention is beneficial may be inclined to report outcome measurements selectively that are favourable to the experimental intervention. Answer 'No' or 'Probably no' if: There is clear evidence (usually through examination of a trial protocol or statistical analysis plan) that all eligible reported results for the outcome domain correspond to all intended outcome measurements. or There is only one possible way in which the outcome domain can be measured (hence there is no opportunity to select from multiple measures). or Outcome measurements are inconsistent across different reports on the same trial, but the trialists have provided the reason for the inconsistency and it is not related to the nature of the results. Answer 'No information' if: Analysis intentions are not available, or the analysis intentions are not reported in sufficient detail to enable an assessment, and there is more than one way in which the outcome domain could have been measured.	
5.3 ... multiple eligible analyses of the data?	A particular outcome measurement may be analysed in multiple ways. Examples include: unadjusted and adjusted models; final value vs change from baseline vs analysis of covariance; transformations of variables; different definitions of composite outcomes (e.g. 'major adverse event'); conversion of continuously scaled outcome to categorical data with different cut-points; different sets of covariates for adjustment; and different strategies for dealing with missing data. Application of multiple methods generates multiple effect estimates for a specific outcome measurement. If multiple estimates are generated but only one or a subset is reported on the basis of the results (e.g. statistical significance), there is a high risk of bias in the fully reported result. Attention should be restricted to analyses that are eligible for consideration by the RoB 2 tool user. For example, if only the result from an analysis of post-intervention values is eligible for inclusion in a meta-analysis (e.g. at 12 weeks after randomization), and this is reported by the trial, then there would not be an issue of selection even if this result was reported (on the basis of the results) in preference to the result from an analysis of changes from baseline. Answer 'Yes' or 'Probably yes' if:	Y/PY/PN/N/NI

	There is clear evidence (usually through examination of a trial protocol or statistical analysis plan) that a measurement was analysed in multiple eligible ways, but data for only one or a subset of analyses is fully reported (without justification), and the fully reported result is likely to have been selected on the basis of the results. Selection on the basis of the results arises from a desire for findings to be newsworthy, sufficiently noteworthy to merit publication, or to confirm a prior hypothesis. For example, trialists who have a preconception or vested interest in showing that an experimental intervention is beneficial may be inclined to selectively report analyses that are favourable to the experimental intervention. Answer 'No' or 'Probably no' if: There is clear evidence (usually through examination of a trial protocol or statistical analysis plan) that all eligible reported results for the outcome measurement correspond to all intended analyses. or There is only one possible way in which the outcome measurement can be analysed (hence there is no opportunity to select from multiple analyses). or Analyses are inconsistent across different reports on the same trial, but the trialists have provided the reason for the inconsistency and it is not related to the nature of the results. Answer 'No information' if: Analysis intentions are not available, or the analysis intentions are not reported in sufficient detail to enable an assessment, and there is more than one way in which the outcome measurement could have been analysed.	
Risk-of-bias judgement	See algorithm.	Low / High / Some concerns
Optional: What is the predicted direction of bias due to selection of the reported result?	If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions.	NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Overall risk of bias

Risk-of-bias judgement		Low / High / Some concerns
Optional: What is the overall predicted direction of bias for this outcome?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable / NA

Overall risk-of-bias judgement	Criteria
Low risk of bias	The study is judged to be at **low risk of bias for all domains** for this result.
Some concerns	The study is judged to raise **some concerns** in at least one domain for this result, but not to be at high risk of bias for any domain.
High risk of bias	The study is judged to be at **high risk of bias** in at least one domain for this result. Or The study is judged to have **some concerns** for **multiple domains** in a way that substantially lowers confidence in the result.

C. ROBINS-I Tool for Non-Randomized Controlled Trials

Risk of bias assessment (cohort-type studies)

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

Bias domain	Signalling questions	Elaboration	Response options
Bias due to confounding	1.1 Is there potential for confounding of the effect of intervention in this study? **If N/PN to 1.1:** the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered	In rare situations, such as when studying harms that are very unlikely to be related to factors that influence treatment decisions, no confounding is expected and the study can be considered to be at low risk of bias due to confounding, equivalent to a fully randomized trial. There is no NI (No information) option for this signalling question.	Y / PY / PN / N
	If Y/PY to 1.1: determine whether there is a need to assess time-varying confounding:		
	1.2. Was the analysis based on splitting participants' follow up time according to intervention received? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, proceed to question 1.3.	If participants could switch between intervention groups then associations between intervention and outcome may be biased by time-varying confounding. This occurs when prognostic factors influence switches between intended interventions.	NA / Y / PY / PN / N / NI
	1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8)	If intervention switches are unrelated to the outcome, for example when the outcome is an unexpected harm, then time-varying confounding will not be present and only control for baseline confounding is required.	NA / Y / PY / PN / N / NI
	Questions relating to baseline confounding only		
	1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?	Appropriate methods to control for measured confounders include stratification, regression, matching, standardization, and inverse probability weighting. They may control for individual variables or for the estimated propensity score. Inverse probability weighting is based on a function of the propensity score. Each method depends on the assumption that there is no unmeasured or residual confounding.	NA / Y / PY / PN / N / NI

	1.5. If Y/PY to 1.4: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Appropriate control of confounding requires that the variables adjusted for are valid and reliable measures of the confounding domains. For some topics, a list of valid and reliable measures of confounding domains will be specified in the review protocol but for others such a list may not be available. Study authors may cite references to support the use of a particular measure. If authors control for confounding variables with no indication of their validity or reliability pay attention to the subjectivity of the measure. Subjective measures (e.g. based on self-report) may have lower validity and reliability than objective measures such as lab findings.	NA / Y / PY / PN / N / NI
	1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?	Controlling for post-intervention variables that are affected by intervention is not appropriate. Controlling for mediating variables estimates the direct effect of intervention and may introduce bias. Controlling for common effects of intervention and outcome introduces bias.	NA / Y / PY / PN / N / NI
	Questions relating to baseline and time-varying confounding		
	1.7. Did the authors use an appropriate analysis method that adjusted for all the important confounding domains and for time-varying confounding?	Adjustment for time-varying confounding is necessary to estimate the effect of starting and adhering to intervention, in both randomized trials and NRSI. Appropriate methods include those based on inverse probability weighting. Standard regression models that include time-updated confounders may be problematic if time-varying confounding is present.	NA / Y / PY / PN / N / NI
	1.8. If Y/PY to 1.7: Were confounding domains that were adjusted for measured validly and reliably by the variables available in this study?	See 1.5 above.	NA / Y / PY / PN / N / NI
	Risk of bias judgement	See Table 1.	Low / Moderate / Serious / Critical / NI
	Optional: What is the predicted direction of bias due to confounding?	Can the true effect estimate be predicted to be greater or less than the estimated effect in the study because one or more of the important confounding domains was not controlled for? Answering this question will be based on expert knowledge and results in other studies and therefore can only be completed after all of the studies in the body of evidence have been reviewed. Consider the potential effect of each of the unmeasured domains and whether all important confounding domains not controlled for in the analysis would be likely to change the estimate in the same direction, or if one important confounding domain that was not controlled for in the analysis is likely to have a dominant impact.	Favours experimental / Favours comparator / Unpredictable

Bias in selection of participants into the study	2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? **If N/PN to 2.1:** go to 2.4	This domain is concerned only with selection into the study based on participant characteristics observed *after* the start of intervention. Selection based on characteristics observed *before* the start of intervention can be addressed by controlling for imbalances between experimental intervention and comparator groups in baseline characteristics that are prognostic for the outcome (baseline confounding).	Y / PY / PN / N / NI
	2.2. If Y/PY to 2.1: Were the post-intervention variables that influenced selection likely to be associated with intervention?	Selection bias occurs when selection is related to an effect of either intervention or a cause of intervention **and** an effect of either the outcome or a cause of the outcome. Therefore, the result is at risk of selection bias if selection into the study is related to both the intervention and the outcome.	NA / Y / PY / PN / N / NI
	2.3 If Y/PY to 2.2: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?		NA / Y / PY / PN / N / NI
	2.4. Do start of follow-up and start of intervention coincide for most participants?	If participants are not followed from the start of the intervention then a period of follow up has been excluded, and individuals who experienced the outcome soon after intervention will be missing from analyses. This problem may occur when prevalent, rather than new (incident), users of the intervention are included in analyses.	Y / PY / PN / N / NI
	2.5. If Y/PY to 2.2 and 2.3, or N/PN to 2.4: Were adjustment techniques used that are likely to correct for the presence of selection biases?	It is in principle possible to correct for selection biases, for example by using inverse probability weights to create a pseudo-population in which the selection bias has been removed, or by modelling the distributions of the missing participants or follow up times and outcome events and including them using missing data methodology. However such methods are rarely used and the answer to this question will usually be "No".	NA / Y / PY / PN / N / NI
	Risk of bias judgement	See Table 1.	Low / Moderate / Serious / Critical / NI
	Optional: What is the predicted direction of bias due to selection of participants into the study?	If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in classification of interventions	3.1 Were intervention groups clearly defined?	A pre-requisite for an appropriate comparison of interventions is that the interventions are well defined. Ambiguity in the definition may lead to bias in the classification of participants. For individual-level interventions, criteria for considering individuals to have received each intervention should be clear and explicit, covering issues such as type, setting, dose, frequency, intensity and/or timing of intervention. For population-level interventions (e.g. measures to control air pollution), the question relates to whether the population is clearly defined, and the answer is likely to be 'Yes'.	Y / PY / PN / N / NI
	3.2 Was the information used to define intervention groups recorded at the start of the intervention?	In general, if information about interventions received is available from sources that could not have been affected by subsequent outcomes, then differential misclassification of intervention status is unlikely. Collection of the information at the time of the intervention makes it easier to avoid such misclassification. For population-level interventions (e.g. measures to control air pollution), the answer to this question is likely to be 'Yes'.	Y / PY / PN / N / NI
	3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?	Collection of the information at the time of the intervention may not be sufficient to avoid bias. The way in which the data are collected for the purposes of the NRSI should also avoid misclassification.	Y / PY / PN / N / NI
	Risk of bias judgement	See Table 1.	Low / Moderate / Serious / Critical / NI
	Optional: What is the predicted direction of bias due to measurement of outcomes or interventions?	If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to deviations from intended interventions	**If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2**		
	4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	Deviations that happen in usual practice following the intervention (for example, cessation of a drug intervention because of acute toxicity) are part of the intended intervention and therefore do not lead to bias in the effect of assignment to intervention. Deviations may arise due to expectations of a difference between intervention and comparator (for example because participants feel unlucky to have been assigned to the comparator group and therefore seek the active intervention, or components of it, or other interventions). Such deviations are not part of usual practice, so may lead to biased effect estimates. However these are not expected in observational studies of individuals in routine care.	Y / PY / PN / N / NI
	4.2. **If Y/PY to 4.1**: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome?	Deviations from intended interventions that do not reflect usual practice will be important if they affect the outcome, but not otherwise. Furthermore, bias will arise only if there is imbalance in the deviations across the two groups.	NA / Y / PY / PN / N / NI
	If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6		
	4.3. Were important co-interventions balanced across intervention groups?	Risk of bias will be higher if unplanned co-interventions were implemented in a way that would bias the estimated effect of intervention. Co-interventions will be important if they affect the outcome, but not otherwise. Bias will arise only if there is imbalance in such co-interventions between the intervention groups. Consider the co-interventions, including any pre-specified co-interventions, that are likely to affect the outcome and to have been administered in this study. Consider whether these co-interventions are balanced between intervention groups.	Y / PY / PN / N / NI
	4.4. Was the intervention implemented successfully for most participants?	Risk of bias will be higher if the intervention was not implemented as intended by, for example, the health care professionals delivering care during the trial. Consider whether implementation of the intervention was successful for most participants.	Y / PY / PN / N / NI
	4.5. Did study participants adhere to the assigned intervention regimen?	Risk of bias will be higher if participants did not adhere to the intervention as intended. Lack of adherence includes imperfect compliance, cessation of intervention, crossovers to the comparator intervention and switches to another active intervention. Consider available information on the proportion of study participants who continued with their assigned	Y / PY / PN / N / NI

		intervention throughout follow up, and answer 'No' or 'Probably No' if this proportion is high enough to raise concerns. Answer 'Yes' for studies of interventions that are administered once, so that imperfect adherence is not possible. We distinguish between analyses where follow-up time after interventions switches (including cessation of intervention) is assigned to (1) the new intervention or (2) the original intervention. (1) is addressed under time-varying confounding, and should not be considered further here.	
	4.6. If N/PN to 4.3, 4.4 or 4.5: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?	It is possible to conduct an analysis that corrects for some types of deviation from the intended intervention. Examples of appropriate analysis strategies include inverse probability weighting or instrumental variable estimation. It is possible that a paper reports such an analysis without reporting information on the deviations from intended intervention, but it would be hard to judge such an analysis to be appropriate in the absence of such information. Specialist advice may be needed to assess studies that used these approaches. If everyone in one group received a co-intervention, adjustments cannot be made to overcome this.	NA / Y / PY / PN / N / NI
	Risk of bias judgement	See Table 2	
	Optional: What is the predicted direction of bias due to deviations from the intended interventions?	If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions.	

Bias due to missing data	5.1 Were outcome data available for all, or nearly all, participants?	"Nearly all" should be interpreted as "enough to be confident of the findings", and a suitable proportion depends on the context. In some situations, availability of data from 95% (or possibly 90%) of the participants may be sufficient, providing that events of interest are reasonably common in both intervention groups. One aspect of this is that review authors would ideally try and locate an analysis plan for the study.	Y / PY / PN / N / NI
	5.2 Were participants excluded due to missing data on intervention status?	Missing intervention status may be a problem. This requires that the *intended* study sample is clear, which it may not be in practice.	Y / PY / PN / N / NI
	5.3 Were participants excluded due to missing data on other variables needed for the analysis?	This question relates particularly to participants excluded from the analysis because of missing information on confounders that were controlled for in the analysis.	Y / PY / PN / N / NI
	5.4 If PN/N to 5.1, or Y/PY to 5.2 or 5.3: Are the proportion of participants and reasons for missing data similar across interventions?	This aims to elicit whether either (i) differential proportion of missing observations or (ii) differences in reasons for missing observations could substantially impact on our ability to answer the question being addressed. "Similar" includes some minor degree of discrepancy across intervention groups as expected by chance.	NA / Y / PY / PN / N / NI
	5.5 If PN/N to 5.1, or Y/PY to 5.2 or 5.3: Is there evidence that results were robust to the presence of missing data?	Evidence for robustness may come from how missing data were handled in the analysis and whether sensitivity analyses were performed by the investigators, or occasionally from additional analyses performed by the systematic reviewers. It is important to assess whether assumptions employed in analyses are clear and plausible. Both content knowledge and statistical expertise will often be required for this. For instance, use of a statistical method such as multiple imputation does not guarantee an appropriate answer. Review authors should seek naïve (complete-case) analyses for comparison, and clear differences between complete-case and multiple imputation-based findings should lead to careful assessment of the validity of the methods used.	NA / Y / PY / PN / N / NI
	Risk of bias judgement	See Table 2	Low / Moderate / Serious / Critical / NI
	Optional: What is the predicted direction of bias due to missing data?	If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in measurement of outcomes	6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	Some outcome measures involve negligible assessor judgment, e.g. all-cause mortality or non-repeatable automated laboratory assessments. Risk of bias due to measurement of these outcomes would be expected to be low.	Y / PY / PN / N / NI
	6.2 Were outcome assessors aware of the intervention received by study participants?	If outcome assessors were blinded to intervention status, the answer to this question would be 'No'. In other situations, outcome assessors may be unaware of the interventions being received by participants despite there being no active blinding by the study investigators; the answer this question would then also be 'No'. In studies where participants report their outcomes themselves, for example in a questionnaire, the outcome assessor is the study participant. In an observational study, the answer to this question will usually be 'Yes' when the participants report their outcomes themselves.	Y / PY / PN / N / NI
	6.3 Were the methods of outcome assessment comparable across intervention groups?	Comparable assessment methods (i.e. data collection) would involve the same outcome detection methods and thresholds, same time point, same definition, and same measurements.	Y / PY / PN / N / NI
	6.4 Were any systematic errors in measurement of the outcome related to intervention received?	This question refers to differential misclassification of outcomes. Systematic errors in measuring the outcome, if present, could cause bias if they are related to intervention or to a confounder of the intervention-outcome relationship. This will usually be due either to outcome assessors being aware of the intervention received or to non-comparability of outcome assessment methods, but there are examples of differential misclassification arising despite these controls being in place.	Y / PY / PN / N / NI
	Risk of bias judgement	See Table 2	Low / Moderate / Serious / Critical / NI
	Optional: What is the predicted direction of bias due to measurement of outcomes?	If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in selection of the reported result	Is the reported effect estimate likely to be selected, on the basis of the results, from... 7.1. ... multiple outcome *measurements* within the outcome domain?	For a specified outcome domain, it is possible to generate multiple effect estimates for different measurements. If multiple measurements were made, but only one or a subset is reported, there is a risk of selective reporting on the basis of results.	Y / PY / PN / N / NI
	7.2 ... multiple *analyses* of the intervention-outcome relationship?	Because of the limitations of using data from non-randomized studies for analyses of effectiveness (need to control confounding, substantial missing data, etc), analysts may implement different analytic methods to address these limitations. Examples include unadjusted and adjusted models; use of final value vs change from baseline vs analysis of covariance; different transformations of variables; a continuously scaled outcome converted to categorical data with different cut-points; different sets of covariates used for adjustment; and different analytic strategies for dealing with missing data. Application of such methods generates multiple estimates of the effect of the intervention versus the comparator on the outcome. If the analyst does not pre-specify the methods to be applied, and multiple estimates are generated but only one or a subset is reported, there is a risk of selective reporting on the basis of results.	Y / PY / PN / N / NI
	7.3 ... different *subgroups*?	Particularly with large cohorts often available from routine data sources, it is possible to generate multiple effect estimates for different subgroups or simply to omit varying proportions of the original cohort. If multiple estimates are generated but only one or a subset is reported, there is a risk of selective reporting on the basis of results.	Y / PY / PN / N / NI
	Risk of bias judgement	See Table 2	Low / Moderate / Serious / Critical / NI
	Optional: What is the predicted direction of bias due to selection of the reported result?	If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Overall bias	**Risk of bias judgement**	See Table 3.	Low / Moderate / Serious / Critical / NI
	Optional: What is the overall predicted direction of bias for this outcome?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Appendix C- GRADE Evidence Profiles

Table C1: GRADE table for vegetables

Question: Should more frequent consumption of vegetables compared to less frequent consumption of vegetables be used for better dietary and health outcomes later in life?

Certainty assessment							Impact	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
Wasting (Assessed with: measurement of body weight using a portable Tanita digital scale with a precision of 0.1kg and measurement of recumbent length using an infant length board with a precision of 0.1cm. Wasting was defined as weight for length z-score less than 2 SD below the mean and was determined using WHO Anthro 2005 v.2.0.4 software).									
1 [a]	observational studies [b]	not serious [c]	serious [d]	not serious [e]	serious [f]	none	Frequency of eating green leafy and orange vegetables was not associated with wasting. 20.5% (n=31), 25.2% (n=31), and 22% (n=26) of those who consumed green leafy and orange vegetables ≥4 times/week, 1-3 times/ week, and never were wasted, respectively (p=0.542 between groups). [g]	⊕◯◯◯ VERY LOW	CRITICAL
Underweight (Assessed with: measurement of body weight using a portable Tanita digital scale with a precision of 0.1kg. Age was determined through interview with the infant's mother. Underweight was defined as a weight for age z-score less than 2 SD below the mean and was determined using WHO Anthro 2005 v.2.0.4 software).									
1 [a]	observational studies [b]	not serious [c]	serious [d]	not serious [e]	serious [f]	none	Frequency of eating green leafy and orange vegetables was not associated with being underweight. 25.8% (n=39), 26% (n=32), and 27.1% (n=32) of those who consumed green leafy and orange vegetables ≥4 times/week, 1-3 times/ week, and never were underweight, respectively (p=0.969 between groups). [g]	⊕◯◯◯ VERY LOW	CRITICAL
Stunting (Assessed with: measurement of recumbent length using an infant length board with a precision of 0.1cm. Age was determined through interview with the infant's mother. Stunting was defined as length for age less than -2 and was determined using WHO Anthro 2005 v.2.0.4 software).									
1 [a]	observational studies [b]	not serious [c]	serious [d]	not serious [e]	serious [f]	none	Frequency of eating green leafy and orange vegetables was not associated with stunting. 29.1% (n=44), 30.1% (n=37), and 23.7% (n=28) of those who consumed green leafy and orange vegetables ≥4 times/week, 1-3 times/ week, and never were stunted, respectively (p=0.491 between groups). [g]	⊕◯◯◯ VERY LOW	CRITICAL

Certainty assessment							Impact	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
Height-for-age Z-score (Follow up: 7 months; Assessed with: measurement of recumbent length to the nearest mm by trained staff using a measuring board. Recumbent length was measured in September 2002, and then again 7 months later during the survey administration. Age was determined using the IPTc trial database of the study area).									
1 [h]	observational studies [i]	serious [j]	serious [k]	not serious [l]	serious [m]	none	In breastfed children aged 9-23 months, those who consumed vegetables/ leaves 0-2 days/ week and ≥ 3 days/week had a mean HAZ of -1.01 (SD=0.93) and -0.58 (SD=0.92), respectively. In age-adjusted models, mean HAZ was -1.01 (p=0.052) and -0.59 (p<0.06), respectively, demonstrating a trend that points to lower HAZ with less vegetable consumption. [n]	⊕◯◯◯ VERY LOW	CRITICAL
Linear Growth (Follow up: 7 months; Assessed with: measurement of recumbent length to the nearest mm by trained staff using a measuring board in September 2002. Followed by a second measurement 7 months later during the survey administration, this time measured as standing height. Height increments were computed as the difference between measurements taken during the survey and in September 2002, divided by the precise duration and multiplied by 7 months (i.e., the average duration of the interval under study). No adjustment was made for the change in measuring technique (I.e., standing height instead of recumbent length).									
1 [h]	observational studies [i]	serious [j]	serious [k]	not serious [l]	serious [m]	none	Frequent consumption of vegetables had an inverse relationship to linear growth (means: 8.3cm and 7.4cm height increment over the preceding 7 months for rare and frequent consumption, respectively, p=0.041). [o]	⊕◯◯◯ VERY LOW	IMPORTANT
Ferritin <20µg/L (Assessed with: venous blood samples that were obtained before 12am and haematological tests were carried out within 6 hours using Technicon H*1. Serum ferritin was analyzed using a immunological method (Orion Hitachi 704)).									
1 [p]	observational studies [q]	very serious [r]	serious [s]	not serious [t]	serious [u]	none	Those who consumed vegetables once/day, compared to <once/day, were significantly more likely to have low iron stores (ferritin values <20µg/L) (Reg coefficient= -2.7, p=0.02). Although the results suggest a negative effect of feeding vegetables once/day, findings were inconsistent as this relationship was not seen for those eating vegetables more frequently (several times/day) compared to <once/day. Additionally, there was no significant relationship observed between the feeding frequency of vegetables and the likelihood of having very low iron stores (ferritin <15µg/L). [v]	⊕◯◯◯ VERY LOW	IMPORTANT

Certainty assessment							Impact	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
Anemia (Assessed with: blood samples obtained through heel or ring finger puncture. Capillary blood was collected in a microcube and the hemoglobin measurement was obtained by direct reading on a portable hemoglobinometer (Hemocue ®). A hemoglobin concentration of 11g/dL was adopted as the cut-off point for the diagnosis of anemia).									
1 [w]	observational studies [x]	very serious [y]	serious [z]	not serious [aa]	serious [bb]	none	Consumption of dark green vegetables was not associated with anemia prevalence. For those who did not consume dark green vegetables in the past 24 hours, compared to those who did, the odds of having anemia were 1.21 times greater (95%CI: 0.67-2.21, p=0.502). 55.6% (n=65) and 60.2% (n=53) of those who consumed dark green vegetables in the past 24 hours did not have anemia, respectively. [cc]	⊕◯◯◯ VERY LOW	CRITICAL
% Stability and Change in Vegetable Consumption (Follow up: 5.5 years; Assessed with: a questionnaire at 3 time points (18 months, 36 months, and 7 years of child age), reported by a parent).									
1 [dd]	observational studies [ee]	very serious [ff]	serious [gg]	not serious [hh]	serious [ii]	none	For boys, overall vegetable consumption at 18 months was positively associated with overall vegetable consumption at 36 months (spearman's rho=0.36) and at 7 years of age (spearman's rho=0.28). For girls, overall vegetable consumption at 18 months was positively associated with overall vegetable consumption at 36 months (spearman's rho=0.37) and at 7 years of age (spearman's rho=0.31). [jj]	⊕◯◯◯ VERY LOW	CRITICAL

Explanations

a. **Ahmad A, Madanijah S, Dwiriani CM, & Kolopaking R. Complementary feeding practices and nutritional status of children 6-23 months old: formative study in Aceh, Indonesia. *Nutrition Research and Practice*. 2018; 12(6): 512-520**. This cross-sectional study took place in 3 sub-districts in Aceh Besar District, Indonesia with 392 healthy children aged 6-23 months.
b. This was a cross-sectional observational study conducted between May-June 2016. Mother's provided detail on frequency of FV, ASF, and NPS consumption through 3 repeated 24hr recalls. Frequency for vegetable consumption was defined as ≥4 times/week, 1-3 times/ week, and never.
c. NIH tool for observational cohort and cross-sectional studies used. Overall, this study has a low risk of selection bias, information bias, and measurement bias and has been rated as good quality for internal validity of cross-sectional studies.
d. Inconsistencies cannot be determined given the lack of data (n=1 study) contributing to this outcome, thus we have downgraded the certainty of evidence for this outcome by 1 level.
e. This study provides direct evidence on population group, exposure, and outcomes of interest.
f. This study did not provide confidence intervals, limiting the ability to make a judgement about imprecision. It should also be noted that the sample size is relatively small (n=392). We have downgraded the certainty of evidence for this outcome by 1 level.
g. Study did not adjust for confounding variables.

h. **Ntab B, Simondon KB, Milet J, Cissé B, Sokhna C, Boulanger D, et al**. **A Young Child Feeding Index Is Not Associated with Either Height-for-Age or Height Velocity in Rural Senegalese Children.** ***The Journal of Nutrition*****. 2004; 135(3): 457-464**. This cross-sectional study took place in rural Senegal with 543 children aged 6-42 months who were included in the Intermittent Preventive Malaria Treatment in children (IPTc) trial.

i. This was a cross-sectional observational study conducted between April-May 2003. Measurements of height and weight were obtained in September 2002 and again 7 months later during survey administration. Height and weight measurements were taken by trained staff using a measuring board and a Seca baby scale, respectively. Caregivers provided detail on frequency of FV and ASF consumption by 24hr recall. Additionally, the number of days each food group had been consumed during the preceding week was also assessed. Frequency was defined as 0-2 days/week and ≥ 3 days/week. A second definition for frequency was characterized as rare and frequent consumption but was never further defined in the paper.

j. NIH tool for observational cohort and cross-sectional studies used. Overall, this study presents some concerns for bias. Firstly, it should be noted that data on food consumption was attained from April to May 2003, while height increment was measured in 2002 and 2003. As food consumed in 2003 does not necessarily mean the child had consumed the same food back in 2002, this may lead to inaccuracies in the association between frequency of food consumption and height increment. Secondly, 7 months may not be long enough to cause a significant change in height. Caution should be taken when interpreting results from this study.

k. Inconsistencies cannot be determined given the lack of data (n=1 study) contributing to this outcome, thus we have downgraded the certainty of evidence for this outcome by 1 level.

l. This study provides direct evidence on population group, exposure, and outcomes of interest.

m. This study did not provide confidence intervals, limiting the ability to make a judgement about imprecision. It should also be noted that the sample size is relatively small (n=543). We have downgraded the certainty of evidence for this outcome by 1 level.

n. Adjusted model by child age.

o. Adjusted model by child age, sex, malaria study intervention group, maternal height, BMI, schooling, and number of children 5 years old.

p. **Wandel M, Fagerli RA, Olsen PT, Borch-Iohnsen B, & Ek J. Iron status and weaning practices among Norwegian and immigrant infants.** ***Nutrition Research*****. 1996; 16(2): 251-265**. This cross-sectional study included 74 healthy infants born between 1 January 1991 and 31 March 1992, who were brought at 1 year of age to Fjell Health Clinic in Drammen, Norway.

q. This was a cross-sectional observational study conducted between January 1992-March 1992. Frequency of FV and ASF consumption were measured using a FFQ through dietary interviews by a primary health care nurse. The FFQ was constructed based on dietary results from a pilot study of a small sample of Norwegian and immigrant children belonging to the same health center. Frequency of consumption of FV was defined as several times per day, once per day, and less than once per day.

r. NIH tool for observational cohort and cross-sectional studies used. This study reported relatively little methodology and what methods are reported are unclear at times, in addition to a very small sample size. Because of this, it's extremely challenging to evaluate whether there was bias introduced or not and ultimately whether the study has internal validity.

s. Inconsistencies cannot be determined given the lack of data (n=1 study) contributing to this outcome, thus we have downgraded the certainty of evidence for this outcome by 1 level.

t. This study provides direct evidence on population group, exposure, and outcomes of interest.

u. This study did not provide confidence intervals, limiting the ability to make a judgement about imprecision. It should also be noted that the sample size is very small (n=74), however the certainty of the evidence has already been rated as very low based on the very poor-quality risk of bias rating and thus could not be further downgraded.

v. Study did not adjust for confounding variables.

w. **Silva DG, Priore SE, & Franceschini SCC. Risk factors for anemia in infants assisted by public health services: the importance of feeding practices and iron supplementation.** ***J Pediatr*****. 2007; 83(2): 149-156**. This cross-sectional study included 205 children aged 6-12 months living in the municipality of Vicosa, Minas Gerais, Brazil.

x. This was a cross-sectional observational study conducted between July 2002-April 2003. Frequency of FV, ASF, and NPS were measured using a FFQ and 24hr recall. Frequency of FV was defined as consumption of food item in previous 24hrs versus no consumption in previous 24hrs. Anemia was defined as hemoglobin <11 g/dL.

y. NIH tool for observational cohort and cross-sectional studies used. Overall, this study has a poor to fair rating. Firstly, there is very limited detail provided on the methodology which makes it challenging to evaluate if there is internal validity. Some considerations to note include lack of sample size justification and no details are provided on participation rate of eligible participants. Furthermore, there is a small sample size with no details provided for baseline characteristics of study sample.

z. Inconsistencies cannot be determined given the lack of data (n=1 study) contributing to this outcome, thus we have downgraded the certainty of evidence for this outcome by 1 level.

aa. This study provides direct evidence on population group, exposure, and outcomes of interest.

bb. The confidence interval around the effect is relatively wide (OR=1.21; 95%CI: 0.67-2.21) and the sample size is small (n=205) thus we have downgraded the certainty of evidence for this outcome by 1 level. However, it should be noted that this study already has a certainty rating of very low based on the very poor-quality risk of bias rating.

cc. Study did not adjust for confounding variables.

dd. **Bjelland M, Brantsaeter AL, Haugen M, Meltzer HM, Nystad W, & Andersen LF. Changes and tracking of fruit, vegetables and sugar-sweetened beverages intake from 18 months to 7 years in the Norwegian mother and child cohort study. *BMC Public Health*. 2013; 13:793**. This longitudinal cohort study took place in a mixture or urban and rural locations in Norway with 9,025 mother infant pairs.

ee. This was a longitudinal cohort study using data from the nation-wide Norwegian Mother and Child Cohort Study, conducted between 1999-2008, with data collection occurring at three time points over 5.5 years (18 months, 36 months, and 7 years of age). Frequency of FV was measured using a questionnaire at the 3 time points. Frequency was categorized into 3 variables: low, medium, and high. For vegetables, low consumption was defined as ≤ 5 times/week, medium was defined as 5.1-7 times/week, and high was defined as >7 times/week.

ff. NIH tool for observational cohort and cross-sectional studies used. Overall, study has a rating of poor and lacks internal validity. Hard to determine whether information bias has occurred due to missing methods and detail in the report. For example, study population, exposure measures, and outcome measures are never clearly defined. Additionally, the study does not report accounting for any confounding variables, and it appears that measurement bias could have occurred. Results should be interpreted with caution.

gg. Inconsistencies cannot be determined given the lack of data (n=1 study) contributing to this outcome, thus we have downgraded the certainty of evidence for this outcome by 1 level.

hh. This study provides direct evidence on population group, exposure, and outcomes of interest.

ii. This study did not provide confidence intervals, limiting the ability to make a judgement about imprecision. It should be noted that the sample size is large (n=9,025 mother infant pairs), however given the lack of confidence intervals provided we have downgraded the certainty of evidence for this outcome by 1 level.

jj. Study did not adjust for confounding variables.

Table C2: GRADE table for fruit

Question: Should more frequent consumption of fruit compared to less frequent consumption of fruit be used for better dietary and health outcomes later in life?

Certainty assessment							Impact	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
Wasting (Assessed with: measurement of body weight using a portable Tanita digital scale with a precision of 0.1kg and measurement of recumbent length using an infant length board with a precision of 0.1cm. Wasting was defined as weight for length z-score less than 2 SD below the mean and was determined using WHO Anthro 2005 v.2.0.4 software).									
1 [a]	observational studies [b]	not serious [c]	serious [d]	not serious [e]	serious [f]	none	Frequency of eating fruits was not associated with wasting. 19.9% (n=28), 26.1% (n=42), and 20% (n=18) of those who consumed fruits ≥3 times/week, 1-2 times/ week, and never were wasted, respectively (p=0.356 between groups). [g]	⊕◯◯◯ VERY LOW	CRITICAL
Underweight (Assessed with: measurement of body weight using a portable Tanita digital scale with a precision of 0.1kg. Age was determined through interview with the infant's mother. Underweight was defined as a weight for age z-score less than 2 SD below the mean and was determined using WHO Anthro 2005 v.2.0.4 software).									
1 [a]	observational studies [b]	not serious [c]	serious [d]	not serious [e]	serious [f]	none	Frequency of eating fruits was not associated with being underweight. 26.2% (n=37), 26.1% (n=42), and 26.7% (n=24) of those who consumed fruits ≥3 times/week, 1-2 times/ week, and never were underweight, respectively (p=0.995 between groups). [g]	⊕◯◯◯ VERY LOW	CRITICAL

Certainty assessment							Impact	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
Stunting (Assessed with: measurement of recumbent length using an infant length board with a precision of 0.1cm. Age was determined through interview with the infant's mother. Stunting was defined as length for age less than -2 and was determined using WHO Anthro 2005 v.2.0.4 software).									
1 [a]	observational studies [b]	not serious [c]	serious [d]	not serious [e]	serious [f]	none	Frequency of eating fruits was not associated with stunting. 30.5% (n=43), 25.5% (n=41), and 27.8% (n=25) of those who consumed fruits ≥3 times/week, 1-2 times/ week, and never were stunted, respectively (p=0.623 between groups). [g]	⊕◯◯◯ VERY LOW	CRITICAL
Height-for-age Z-score (Follow up: 7 months; Assessed with: measurement of recumbent length to the nearest mm by trained staff using a measuring board. Recumbent length was measured in September 2002, and then again 7 months later during the survey administration. Age was determined using the IPTc trial database of the study area).									
1 [h]	observational studies [i]	serious [j]	serious [k]	not serious [l]	serious [m]	none	In breastfed children aged 9-23 months, those who consumed fruit 0-2 days/week and ≥ 3 days/week had a mean HAZ of -1.02 (SD=0.93) and -0.75 (SD=0.94), respectively. In age-adjusted models, mean HAZ was -1.04 (p=0.051) and -0.71 (p=0.059), respectively, demonstrating a trend that points to lower HAZ with less fruit consumption. [n]	⊕◯◯◯ VERY LOW	CRITICAL
Linear Growth (Follow up: 7 months; Assessed with: measurement of recumbent length to the nearest mm by trained staff using a measuring board in September 2002. Followed by a second measurement 7 months later during the survey administration, this time measured as standing height. Height increments were computed as the difference between measurements taken during the survey and in September 2002, divided by the precise duration and multiplied by 7 months (I.e., the average duration of the interval under study). No adjustment was made for the change in measuring technique (I.e., standing height instead of recumbent length).									
1 [h]	observational studies [i]	serious [j]	serious [k]	not serious [l]	serious [m]	none	Fruit consumption was positively associated with linear growth in fully adjusted models (means: 7.9cm and 8.7cm height increment over the preceding 7 months for rare and frequent consumption, respectively, p=0.027). [o]	⊕◯◯◯ VERY LOW	IMPORTANT
Anemia (Assessed with: blood samples obtained through heel or ring finger puncture. Capillary blood was collected in a microcube and the hemoglobin measurement was obtained by direct reading on a portable hemoglobinometer (Hemocue ®). A hemoglobin concentration of 11g/dL was adopted as the cut-off point for the diagnosis of anemia).									

Certainty assessment							Impact	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
1 [p]	observational studies [q]	very serious [r]	serious [s]	not serious [t]	serious [u]	none	Fruit consumption was not significantly associated with anemia prevalence when considering consumption in the past 24 hours. For those who did not consume fruit in the past 24 hours, compared to those who did, the odds of having anemia were 1.24 times greater (95% CI: 0.25-6.75, p=0.537). 57.4% (n=113) and 62.5% (n=5) of those who consumed fruit in the past 24 hours and not, had anemia, respectively. When looking at daily vs < than daily frequencies, fruit consumption was significantly associated with anemia prevalence. In the adjusted model, those who consumed fruit < daily, compared to daily, had 1.88 (95% CI: 1.03-3.42, p=0.003) greater odds of having anemia. 48.7% (n=55) and 68.5% (n=63) of those who consumed fruit daily and < daily had anemia, respectively. [v]	⊕◯◯◯ VERY LOW	CRITICAL
% Stability and Change in Fruit Consumption (Follow up: 5.5 years; Assessed with: a questionnaire at 3 time points (18 months, 36 months, and 7 years of child age), reported by a parent).									
1 [w]	observational studies [x]	very serious [y]	serious [z]	not serious [aa]	serious [bb]	none	For boys, overall fruit consumption at 18 months was positively associated with overall fruit consumption at 36 months (spearman's rho=0.36) and at 7 years of age (spearman's rho=0.23). For girls, overall fruit consumption at 18 months was positively associated with overall fruit consumption at 36 months (spearman's rho=0.36) and at 7 years of age (spearman's rho=0.24). [cc]	⊕◯◯◯ VERY LOW	CRITICAL

Explanations

a. **Ahmad A, Madanijah S, Dwiriani CM, & Kolopaking R. Complementary feeding practices and nutritional status of children 6-23 months old: formative study in Aceh, Indonesia. *Nutrition Research and Practice*. 2018; 12(6): 512-520**. This cross-sectional study took place in 3 sub-districts in Aceh Besar District, Indonesia with 392 healthy children aged 6-23 months.

b. This was a cross-sectional observational study conducted between May-June 2016. Mother's provided detail on frequency of FV, ASF, and NPS consumption through 3 repeated 24hr recalls. Frequency for fruit consumption was defined as ≥3 times/week, 1-2 times/ week, and never.

c. NIH tool for observational cohort and cross-sectional studies used. Overall, this study has a low risk of selection bias, information bias, and measurement bias and has been rated as good quality for internal validity of cross-sectional studies.

d. Inconsistencies cannot be determined given the lack of data (n=1 study) contributing to this outcome, thus we have downgraded the certainty of evidence for this outcome by 1 level.
e. This study provides direct evidence on population group, exposure, and outcomes of interest.
f. This study did not provide confidence intervals, limiting the ability to make a judgement about imprecision. It should also be noted that the sample size is relatively small (n=392). We have downgraded the certainty of evidence for this outcome by 1 level.
g. Study did not adjust for confounding variables.
h. **Ntab B, Simondon KB, Milet J, Cissé B, Sokhna C, Boulanger D, et al. A Young Child Feeding Index Is Not Associated with Either Height-for-Age or Height Velocity in Rural Senegalese Children. *The Journal of Nutrition*. 2004; 135(3): 457-464**. This cross-sectional study took place in rural Senegal with 543 children aged 6-42 months who were included in the Intermittent Preventive Malaria Treatment in children (IPTc) trial.
i. This was a cross-sectional observational study conducted between April-May 2003. Measurements of height and weight were obtained in September 2002 and again 7 months later during survey administration. Height and weight measurements were taken by trained staff using a measuring board and a Seca baby scale, respectively. Caregivers provided detail on frequency of FV and ASF consumption by 24hr recall. Additionally, the number of days each food group had been consumed during the preceding week was also assessed. Frequency was defined as 0-2 days/week and ≥ 3 days/week. A second definition for frequency was characterized as rare and frequent consumption but was never further defined in the paper.
j. NIH tool for observational cohort and cross-sectional studies used. Overall, this study presents some concerns for bias. Firstly, it should be noted that data on food consumption was attained from April to May 2003, while height increment was measured in 2002 and 2003. As food consumed in 2003 does not necessarily mean the child had consumed the same food back in 2002, this may lead to inaccuracies in the association between frequency of food consumption and height increment. Secondly, 7 months may not be long enough to cause a significant change in height. Caution should be taken when interpreting results from this study.
k. Inconsistencies cannot be determined given the lack of data (n=1 study) contributing to this outcome, thus we have downgraded the certainty of evidence for this outcome by 1 level.
l. This study provides direct evidence on population group, exposure, and outcomes of interest.
m. This study did not provide confidence intervals, limiting the ability to make a judgement about imprecision. It should also be noted that the sample size is relatively small (n=543). We have downgraded the certainty of evidence for this outcome by 1 level.
n. Adjusted model by child age.
o. Adjusted model by child age, sex, malaria study intervention group, maternal height, BMI, schooling, and number of children 5 years old.
p. **Silva DG, Priore SE, & Franceschini SCC. Risk factors for anemia in infants assisted by public health services: the importance of feeding practices and iron supplementation. *J Pediatr*. 2007; 83(2): 149-156**. This cross-sectional study included 205 children aged 6-12 months living in the municipality of Vicosa, Minas Gerais, Brazil.
q. This was a cross-sectional observational study conducted between July 2002-April 2003. Frequency of FV, ASF, and NPS were measured using a FFQ and 24hr recall. Frequency of FV was defined as consumption of food item in previous 24hrs versus no consumption in previous 24hrs. Additionally for fruit, frequency was defined as consumed daily versus less than daily. Anemia was defined as hemoglobin <11 g/dL.
r. NIH tool for observational cohort and cross-sectional studies used. Overall, this study has a poor to fair rating. Firstly, there is very limited detail provided on the methodology which makes it challenging to evaluate if there is internal validity. Some considerations to note include lack of sample size justification and no details are provided on participation rate of eligible participants. Furthermore, there is a small sample size with no details provided for baseline characteristics of study sample.
s. Inconsistencies cannot be determined given the lack of data (n=1 study) contributing to this outcome, thus we have downgraded the certainty of evidence for this outcome by 1 level.
t. This study provides direct evidence on population group, exposure, and outcomes of interest.
u. The confidence interval around the effect is relatively wide (OR=1.88; 95%CI: 1.03-3.42) and the sample size is small (n=205) thus we have downgraded the certainty of evidence for this outcome by 1 level. However, it should be noted that this study already has a certainty rating of very low based on the very poor-quality risk of bias rating.
v. Adjusted model by family income per capita and consumption of medicated iron supplements.
w. **Bjelland M, Brantsaeter AL, Haugen M, Meltzer HM, Nystad W, & Andersen LF. Changes and tracking of fruit, vegetables and sugar-sweetened beverages intake from 18 months to 7 years in the Norwegian mother and child cohort study. *BMC Public Health*. 2013; 13:793**. This longitudinal cohort study took place in a mixture or urban and rural locations in Norway with 9,025 mother infant pairs.
x. This was a longitudinal cohort study using data from the nation-wide Norwegian Mother and Child Cohort Study, conducted between 1999-2008, with data collection occurring at three time points over 5.5 years (18 months, 36 months, and 7 years of age). Frequency of FV was measured using a questionnaire at the 3 time points. Frequency was categorized into 3 variables: low, medium, and high. For fruits, low consumption was defined as ≤ 5 times/week, medium was defined as 5.1-13.9 times/week, and high was defined as >14 times/week.
y. NIH tool for observational cohort and cross-sectional studies used. Overall, study has a rating of poor and lacks internal validity. Hard to determine whether information bias has occurred due to missing methods and detail in the report. For example, study population, exposure measures, and outcome measures are never clearly defined.

Additionally, the study does not report accounting for any confounding variables, and it appears that measurement bias could have occurred. Results should be interpreted with caution.

z. Inconsistencies cannot be determined given the lack of data (n=1 study) contributing to this outcome, thus we have downgraded the certainty of evidence for this outcome by 1 level.

aa. This study provides direct evidence on population group, exposure, and outcomes of interest.

bb. This study did not provide confidence intervals, limiting the ability to make a judgement about imprecision. It should be noted that the sample size is large (n=9,025 mother infant pairs), however given the lack of confidence intervals provided we have downgraded the certainty of evidence for this outcome by 1 level.

cc. Study did not adjust for confounding variables.

Table C3: GRADE table for legumes and beans

Question: Should more frequent consumption of legumes and beans compared to less frequent consumption of legumes and beans be used for better dietary and health outcomes later in life?

Certainty assessment							Impact	Certainty	Importance
№ of studies	**Study design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**			
Wasting (Assessed with: measurement of body weight using a portable Tanita digital scale with a precision of 0.1kg and measurement of recumbent length using an infant length board with a precision of 0.1cm. Wasting was defined as weight for length z-score less than 2 SD below the mean and was determined using WHO Anthro 2005 v.2.0.4 software).									
1 [a]	observational studies [b]	not serious [c]	serious [d]	not serious [e]	serious [f]	none	Frequency of eating legumes was not associated with wasting. 25.9% (n=15), 24% (n=37), and 20% (n=36) of those who consumed legumes ≥3 times/week, 1-2 times/ week, and never were wasted, respectively (p=0.542 between groups). [g]	⊕◯◯◯ VERY LOW	CRITICAL
Underweight (Assessed with: measurement of body weight using a portable Tanita digital scale with a precision of 0.1kg. Age was determined through interview with the infant's mother. Underweight was defined as a weight for age z-score less than 2 SD below the mean and was determined using WHO Anthro 2005 v.2.0.4 software).									
1 [a]	observational studies [b]	not serious [c]	serious [d]	not serious [e]	serious [f]	none	Frequency of eating legumes was not associated with being underweight. 36.2% (n=21), 25.3% (n=39), and 23.9% (n=43) of those who consumed legumes ≥3 times/week, 1-2 times/ week, and never were underweight, respectively (p=0.174 between groups). [g]	⊕◯◯◯ VERY LOW	CRITICAL
Stunting (Assessed with: measurement of recumbent length using an infant length board with a precision of 0.1cm. Age was determined through interview with the infant's mother. Stunting was defined as length for age less than -2 and was determined using WHO Anthro 2005 v.2.0.4 software).									
1 [a]	observational studies [b]	not serious [c]	serious [d]	not serious [e]	serious [f]	none	Frequency of eating legumes was not associated with stunting. 32.8% (n=19), 27.9% (n=43), and 26.1% (n=47) of those who consumed legumes ≥3 times/week, 1-2 times/ week, and never were stunted, respectively (p=0.618 between groups). [g]	⊕◯◯◯ VERY LOW	CRITICAL

Certainty assessment							Impact	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
Anemia (Assessed with: blood samples obtained through heel or ring finger puncture. Capillary blood was collected in a microcube and the hemoglobin measurement was obtained by direct reading on a portable hemoglobinometer (Hemocue ®). A hemoglobin concentration of 11g/dL was adopted as the cut-off point for the diagnosis of anemia).									
1 [h]	observational studies [i]	very serious [j]	serious [k]	not serious [l]	serious [m]	none	Consumption of beans was not significantly associated with anemia prevalence. For those who consumed beans < daily, compared to daily, the odds of having anemia were 0.8 times less (95% CI: 0.36-1.78, p=0.550). 58.5% (n=100) and 52.9% (n=18) of those who consumed beans daily and <daily had anemia, respectively. [n]	⊕◯◯◯ VERY LOW	CRITICAL

Explanations

a. **Ahmad A, Madanijah S, Dwiriani CM, & Kolopaking R. Complementary feeding practices and nutritional status of children 6-23 months old: formative study in Aceh, Indonesia. *Nutrition Research and Practice*. 2018; 12(6): 512-520**. This cross-sectional study took place in 3 sub-districts in Aceh Besar District, Indonesia with 392 healthy children aged 6-23 months.
b. This was a cross-sectional observational study conducted between May-June 2016. Mother's provided detail on frequency of FV, ASF, and NPS consumption through 3 repeated 24hr recalls. Frequency for NPS consumption was defined as ≥3 times/week, 1-2 times/ week, and never.
c. NIH tool for observational cohort and cross-sectional studies used. Overall, this study has a low risk of selection bias, information bias, and measurement bias and has been rated as good quality for internal validity of cross-sectional studies.
d. Inconsistencies cannot be determined given the lack of data (n=1 study) contributing to this outcome, thus we have downgraded the certainty of evidence for this outcome by 1 level.
e. This study provides direct evidence on population group, exposure, and outcomes of interest.
f. This study did not provide confidence intervals, limiting the ability to make a judgement about imprecision. It should also be noted that the sample size is relatively small (n=392). We have downgraded the certainty of evidence for this outcome by 1 level.
g. Study did not adjust for confounding variables.
h. **Silva DG, Priore SE, & Franceschini SCC. Risk factors for anemia in infants assisted by public health services: the importance of feeding practices and iron supplementation. *J Pediatr*. 2007; 83(2): 149-156**. This cross-sectional study included 205 children aged 6-12 months living in the municipality of Vicosa, Minas Gerais, Brazil.
i. This was a cross-sectional observational study conducted between July 2002-April 2003. Frequency of FV, ASF, and NPS were measured using a FFQ and 24hr recall. Frequency of NPS consumption was defined as consumed daily versus less than daily. Anemia was defined as hemoglobin <11 g/dL.
j. NIH tool for observational cohort and cross-sectional studies used. Overall, this study has a poor to fair rating. Firstly, there is very limited detail provided on the methodology which makes it challenging to evaluate if there is internal validity. Some considerations to note include lack of sample size justification and no details are provided on participation rate of eligible participants. Furthermore, there is a small sample size with no details provided for baseline characteristics of study sample.
k. Inconsistencies cannot be determined given the lack of data (n=1 study) contributing to this outcome, thus we have downgraded the certainty of evidence for this outcome by 1 level.
l. This study provides direct evidence on population group, exposure, and outcomes of interest.
m. The confidence interval around the effect is relatively wide (OR=0.80; 95%CI: 0.36-1,78) and the sample size is small (n=205) thus we have downgraded the certainty of evidence for this outcome by 1 level. However, it should be noted that this study already has a certainty rating of very low based on the very poor-quality risk of bias rating.
n. Study did not adjust for confounding variables.

Table C4: GRADE table for fruit and vegetables combined

Question: Should more frequent consumption of fruit and vegetables compared to less frequent consumption of fruit and vegetables be used for better dietary and health outcomes later in life?

Certainty assessment							Impact	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
Stunting (Assessed with: the China Food and Nutrition Surveillance System (CFNSS). The CFNSS uses the Z-score as recommended by WHO to evaluate the growth and development of infants and young children).									
1 [a]	observational studies [b]	very serious [c]	serious [d]	not serious [e]	serious [f]	none	Frequency of eating vegetables and fruits was associated with stunting prevalence. For those who consumed vegetables and fruits weekly, monthly, and < once per month or none, compared to daily, the odds of being stunted were 1.739 (p=0.00), 1.698 (p=0.03), and 1.768 (p=0.00) times greater, respectively. [g]	⨁◯◯◯ VERY LOW	CRITICAL
Underweight (Assessed with: the CFNSS which uses the Z-score as recommended by WHO to evaluate the growth and development of infants and young children).									
1 [a]	observational studies [b]	very serious [c]	serious [d]	not serious [e]	serious [f]	none	Frequency of eating vegetables and fruits was associated with underweight prevalence. For those who consumed vegetables and fruits weekly, monthly, and < once per month or none, compared to daily, the odds of being underweight were 1.908 (p=0.00), 1.566 (p=0.10), and 1.478 (p=0.01) times greater, respectively. [g]	⨁◯◯◯ VERY LOW	CRITICAL

Explanations

a. **Chang S, Chen C, He W, & Wang Y. Analysis on the changes of nutritional status among Chinese infants and young children -- the improvement of complementary feeding. *Journal of Hygiene Research.* 2007; 36(2):207-9**. This cross-sectional study took place in a mixture of urban and rural locations in China with 13,107 infants aged 6-24 months old.
b. This was a cross-sectional observational study using data from the China Food and Nutrition Surveillance System (CFNSS). Data collection occurred between 1992-2005. Frequency of FV consumption was defined as daily, weekly, monthly, less than once per month, or none.
c. NIH tool for observational cohort and cross-sectional studies used. This study lacks internal validity. Firstly, the paper does not report a methods section. Thus, no detail is provided on the tools used to measure the exposure or outcome. The sample size is not clearly stated, and thus had to be inferred from tables. In the analysis, the reference arm is not reported within the tables, once again requiring the reader to infer this information from the text. Results should be interpreted with extreme caution.
d. Inconsistencies cannot be determined given the lack of data (n=1 study) contributing to this outcome, thus we have downgraded the certainty of evidence for this outcome by 1 level.
e. This study provides direct evidence on population group, exposure, and outcomes of interest.
f. This study did not provide confidence intervals, limiting the ability to make a judgement about imprecision. It should be noted that the sample size is very large (n=13,107), however given the lack of confidence intervals provided we have downgraded the certainty of evidence for this outcome by 1 level.
g. Study did not adjust for confounding variables.

Appendix D- Summary of Study Demographics Table

Author	Year Published	Study Setting			Study Design *	Age Range (months)	Sample Size	Food Groups*			Quality Assessment
		Country	WHO Region	Urban/ Rural				F/V	NPS	ASF	
Ahmad, A. et al	2018	Indonesia	South-East Asia	Not reported	CS	6-23	392	✓	✓	✓	Good
Bjelland, M. et al	2013	Norway	Europe	Mix	Cohort	18 months-7 yrs	9,940	✓			Poor
Ntab, B. et al	2005	Senegal	Africa	Rural	CS	6-42	543	✓		✓	Fair
Silva, D.G. et al	2007	Brazil	Americas	Urban	CS	6-12	205	✓	✓	✓	Poor
Wandel, M. et al	1996	Norway	Europe	Urban	CS	12 months	90	✓		✓	Poor
Chang, S. et al	2007	China	Western Pacific	Mix	CS	6-24	13, 107	✓		✓	Poor

*Study Design: = cross-sectional – CS; Food Groups: F/V – Fruits & vegetables; NSP – Nuts, pulses, seeds; ASF – Animal-source foods

Appendix E- Detailed Study Characteristics of Included Studies Tables

Ahmad 2018

Methods	This was a cross-sectional study conducted in May-June 2016 in 3 sub-districts in Aceh Besar District, Indonesia.
Participants	392 children aged 6-23 months where the mother agreed to be interviewed, the child was healthy, not experiencing chronic/ congenital diseases, such as heart abnormality, and not having experienced acute conditions such as fever, diarrhea, and respiratory infections in the previous 2 weeks.
Interventions	Mother's provided detail on frequency of FV, ASF, and NPS consumption through 3 repeated 24hr recalls. Frequency for vegetable consumption was defined as ≥4 times/week, 1-3 times/ week, and never. For fruit and NPS consumption, frequency was defined as ≥3 times/week, 1-2 times/ week, and never.
Outcomes	Wasting, underweight, and stunting.

Ntab 2005

Methods	This was a cross-sectional study conducted in rural Senegal. Food consumption and anthropometric survey was conducted in April-May 2003. Measurements of height and weight were obtained in September 2002.

Participants	543 children who were included in the Intermittent Preventive Malaria Treatment in children (IPTc) trial, aged between 6-42 months, residing in 1 of the 10 villages at the time of study, and whose mother provided oral informed consent (or by another primary caregiver).
Interventions	Caregiver's provided detail on frequency of FV and ASF consumption by 24hr recall. Additionally, the number of days each food group had been consumed during the preceding week was also assessed. Frequency was defined as 0-2 days/week and ≥ 3 days/week. A second definition for frequency was characterized as rare and frequent consumption but was never further defined in the paper.
Outcomes	HAZ and linear growth.

Wandel 1996

Methods	This was a cross-sectional study conducted in urban Norway from January 1991 to March 1992.
Participants	74 healthy infants born between 1 January 1991 and 31 March 1992, who were brought to health control at 1 year of age at Fjell Health Clinic in Drammen, Norway.
Interventions	Frequency of FV and ASF consumption were measured using a FFQ through dietary interviews by a primary health care nurse. The FFQ was constructed based on dietary results from a pilot study of a small sample of Norwegian and immigrant children belonging to the same health center. Frequency of FV was defined as several per day, once per day, and less than once per day.
Outcomes	Iron status based on serum ferritin levels.

Silva 2007

Methods	This was a cross-sectional study conducted in urban Brazil from July 2002 to April 2003.

Participants	205 children aged 6-12 months living in the municipality of Viçosa, located in the state of Minas Gerais.
Interventions	Frequency of FV, ASF, and NPS were measured using a FFQ and 24hr recall. Frequency of FV was defined as consumption of food item in previous 24hrs versus no consumption in previous 24hrs. Additionally for fruit and for NPS, frequency was defined as consumed daily versus less than daily.
Outcomes	Anemia defined as hemoglobin <11 g/dL.

Bjelland 2013

Methods	This was a longitudinal study using data from the nation-wide Norwegian Mother and Child Cohort Study. This study was conducted in a mixture of urban and rural locations in Norway from 1999 to 2008, with data collection occurring at three time points over 5.5 years (18 months, 36 months, and 7yrs of age).
Participants	Any pregnant woman who was recruited between the data collection timeframe was eligible for the study. Of all women, 38.5% consented to participate leaving 9,025 mother infant pairs.
Interventions	Frequency of FV was measured using a questionnaire at the 3 time points. Frequency was categorized into 3 variables: low, medium, and high. For vegetables, low consumption was defined as ≤ 5 times/week, medium was defined as 5.1-7 times/week, and high was defined as >7 times/week. For fruits, low consumption was defined as ≤ 5 times/week, medium was defined as 5.1-13.9 times/week, and high was defined as >14 times/week.
Outcomes	Changes and tracking in intake of fruit and vegetables at 36 months and 7 years of age.

Chang 2007

Methods	This was a cross-sectional study using data from the China Food and Nutrition Surveillance System (CFNSS). This study was conducted in a mixture of urban and rural locations in China. Data collection occurred from 1992 to 2005.
Participants	13,107 infants aged 6-24 months old.
Interventions	Frequency of FV was measured through the CFNSS survey and defined as daily, weekly, monthly, less than once per month, or none.
Outcomes	Stunting and underweight.

Appendix F- Risk of Bias of Included Studies Tables

Ahmad 2018

Criteria	Authors' Judgement	Criteria	Authors' Judgement
Was the research question or objective in this paper clearly stated?	Yes	For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (e.g., categories of exposure, or exposure measured as a continuous variable)?	Yes
Was the study population clearly specified and defined?	Yes	Were the exposure measures clearly defined, valid, reliable, and implemented consistently across all study participants?	Yes
Was the participation rate of eligible persons at least 50%?	Yes	Was the exposure(s) assessed more than once over time?	Yes
Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion criteria for being in the study prescribed and applied uniformly to all participants?	Yes	Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?	Yes

Was a sample size justification, power description, or variance and effect estimates provided?	Yes	Were the outcome assessors blinded to the exposure status of participants?	Not reported
For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured?	Cannot determine	Was loss to follow up after baseline 20% or less?	Yes
Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed?	Not applicable	Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcome(s)?	Cannot determine
Overall Quality Rating	Good		
Additional Comments	Overall, study is well controlled and presents a low risk for bias in cross-sectional studies.		

Ntab 2005

Criteria	Authors' Judgement	Criteria	Authors' Judgement
Was the research question or objective in this paper clearly stated?	Yes	For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (e.g., categories of exposure, or exposure measured as a continuous variable)?	Yes

Was the study population clearly specified and defined?	Yes	Were the exposure measures clearly defined, valid, reliable, and implemented consistently across all study participants?	Yes
Was the participation rate of eligible persons at least 50%?	Yes	Was the exposure(s) assessed more than once over time?	No
Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion criteria for being in the study prescribed and applied uniformly to all participants?	Yes	Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?	Yes
Was a sample size justification, power description, or variance and effect estimates provided?	No	Were the outcome assessors blinded to the exposure status of participants?	No
For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured?	No	Was loss to follow up after baseline 20% or less?	Yes
Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if existed?	Cannot determine	Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcome(s)?	Yes

Overall Quality Rating	Fair
Additional Comments	Overall, this study presents some concerns for bias. Firstly, it should be noted that data on food consumption was attained from April to May 2003, while height increment was measured in 2002 and 2003. As food consumed in 2003 does not necessarily mean the child had consumed the same food back in 2002, this may lead to inaccuracies in the association between frequency of food consumption and height increment. Secondly, 7 months may not be long enough to cause a significant change in height. Caution should be taken when reading results from this study.

Wandel 1996

Criteria	Authors' Judgement	Criteria	Authors' Judgement
Was the research question or objective in this paper clearly stated?	Yes	For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (e.g., categories of exposure, or exposure measured as a continuous variable)?	Yes
Was the study population clearly specified and defined?	No	Were the exposure measures clearly defined, valid, reliable, and implemented consistently across all study participants?	Yes
Was the participation rate of eligible persons at least 50%?	Yes	Was the exposure(s) assessed more than once over time?	No

Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion criteria for being in the study prescribed and applied uniformly to all participants?	Yes	Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?	Yes
Was a sample size justification, power description, or variance and effect estimates provided?	No	Were the outcome assessors blinded to the exposure status of participants?	No
For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured?	No	Was loss to follow up after baseline 20% or less?	Yes
Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if existed?	Not applicable	Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcome(s)?	Cannot determine
Overall Quality Rating	Poor		
Additional Comments	This study reported relatively little methodology and what methods are reported are unclear at times, in addition to a very small sample size. Because of this, it's extremely challenging to evaluate whether there was bias introduced or not and ultimately whether the study has internal validity.		

Silva 2007

Criteria	Authors' Judgement	Criteria	Authors' Judgement
Was the research question or objective in this paper clearly stated?	Yes	For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (e.g., categories of exposure, or exposure measured as a continuous variable)?	Yes
Was the study population clearly specified and defined?	No	Were the exposure measures clearly defined, valid, reliable, and implemented consistently across all study participants?	Not reported
Was the participation rate of eligible persons at least 50%?	Cannot determine	Was the exposure(s) assessed more than once over time?	No
Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion criteria for being in the study prescribed and applied uniformly to all participants?	Yes	Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?	Yes
Was a sample size justification, power description, or variance and effect estimates provided?	No	Were the outcome assessors blinded to the exposure status of participants?	No

For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured?	No	Was loss to follow up after baseline 20% or less?	Yes
Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if existed?	Not applicable	Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcome(s)?	Yes
Overall Quality Rating	Poor		
Additional Comments	Overall, this study has a poor to fair rating. Firstly, there is very limited detail provided on the methodology which makes it challenging to evaluate if there is internal validity. Some considerations to note include lack of sample size justification, no mention of loss-to-follow up, or details provided on participation rate of eligible participants. Furthermore, there is a small sample size with no details provided for baseline characteristics of study sample.		

Bjelland 2013

Criteria	Authors' Judgement	Criteria	Authors' Judgement
Was the research question or objective in this paper clearly stated?	Yes	For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (e.g., categories of exposure, or exposure measured as a continuous variable)?	Yes

Was the study population clearly specified and defined?	No	Were the exposure measures clearly defined, valid, reliable, and implemented consistently across all study participants?	No
Was the participation rate of eligible persons at least 50%?	No	Was the exposure(s) assessed more than once over time?	Yes
Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion criteria for being in the study prescribed and applied uniformly to all participants?	Cannot determine	Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?	No
Was a sample size justification, power description, or variance and effect estimates provided?	No	Were the outcome assessors blinded to the exposure status of participants?	No
For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured?	Yes	Was loss to follow up after baseline 20% or less?	No
Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if existed?	Yes	Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcome(s)?	No

Overall Quality Rating	Poor
Additional Comments	Overall, study has a rating of poor and lacks internal validity. Hard to determine whether information bias has occurred due to missing methods and detail in the report. For example, study population, exposure measures, and outcome measures are never clearly defined. Additionally, the study does not report accounting for any confounding variables, and it appears that measurement bias could have occurred. Results should be interpreted with caution.

Chang 2007

Criteria	Authors' Judgement	Criteria	Authors' Judgement
Was the research question or objective in this paper clearly stated?	Yes	For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (e.g., categories of exposure, or exposure measured as a continuous variable)?	Yes
Was the study population clearly specified and defined?	No	Were the exposure measures clearly defined, valid, reliable, and implemented consistently across all study participants?	No
Was the participation rate of eligible persons at least 50%?	Cannot determine	Was the exposure(s) assessed more than once over time?	No

Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion criteria for being in the study prescribed and applied uniformly to all participants?	Cannot determine	Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?	No
Was a sample size justification, power description, or variance and effect estimates provided?	No	Were the outcome assessors blinded to the exposure status of participants?	Cannot determine
For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured?	Cannot determine	Was loss to follow up after baseline 20% or less?	Cannot determine
Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed?	Not applicable	Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcome(s)?	Cannot determine
Overall Quality Rating	Poor		

Additional Comments	This study lacks internal validity. Firstly, the paper does not report a methods section. Thus, no detail is provided on tools used to measure the exposure or outcome. The sample size is not clearly stated, and thus had to be inferred from tables. In the analysis, the reference arm is not reported within the tables, once again requiring the reader to infer this information from the text. Results should be interpreted with extreme caution.