CHAPTER 2: REVIEW OF LITERATURE

Chapter overview

2.1 NUTRITION OF PRIMARY SCHOOL CHILDREN

There is general consensus that appropriate eating habits of children are important for optimal physical and cognitive development, the attainment of healthy weight and the reduction of the risk of CNCD.[16] In order to achieve nutritional health many organisations and countries have set up dietary guidelines, including South Africa, where country-specific, evidence-based food-based guidelines for people seven years and older were officially approved and adopted by government in 2003.[14]

In 2002 the Institute of Medicine's (IOM) Food and Nutrition Board released the Dietary Reference Intakes (DRI) for energy, carbohydrates, fat, fatty acids, and cholesterol, thereby updating the Recommended Dietary Allowances (RDA). Acceptable macronutrient distribution ranges (AMDR) as a percent of energy intake for carbohydrate, fat, and protein for children are as follows: [10, 16]

- Carbohydrate: 45% to 65% of total energy
- Fat: 30% to 40% of energy for 1 to 3 years
  25% to 35% of energy for 4 to 19 years
- Protein: 5%- 20% for young children
  10% to 30% for older children

Knowledge of current eating habits of children relative to these dietary guidelines and reference intakes is an important starting point for appropriate intervention for maintenance or improvement of the nutritional status of children. In the following section some major findings of relevance to the current research context are presented, first from the international literature and then from South Africa.

2.1.1 Eating habits of children

Dietary intake data of children on the international arena suggests the following:

Amongst eleven to 18-year-old United States (US) adolescents from 1965 to 1996 total energy intakes decreased, as did the proportion of energy from fat (from 39% to 32%) and saturated fat (15% to 12%). There were concurrent increases in consumption of higher fat potato and mixed dishes (pizza and macaroni cheese), lower fat milk replaced higher fat milks, but total milk consumption decreased by 36%.[51] Ten-year results of the NHLBI Growth and Health Study showed that for girls total and saturated fat and cholesterol intakes had decreased with age. In all cases the decrease was more in white girls than in black girls. A substantial percentage of both ethnic groups had not yet reached NCEP goals in terms of PFE, PSFE and cholesterol.[52] Even children known to be at high risk for cardiovascular disease (based on the NCEP criteria) were no more likely to meet guidelines for heart-healthy diets than were children at low risk.[53]

Dwyer et al [54] investigated the eating patterns of US adolescents and found that an increase in eating occasions was common. It was associated with increased energy intake but a reduced relative amount of total and saturated fat consumed. 'Grazing' may be the modal behaviour of children who also increasingly make their own food decisions.[55] A high percentage of daily food intakes of children occurs at schools. In the US, school stores were found to sell primarily snacks with high fat and sugar content.[56] Amongst participants in the Bogalusa Heart Study, Nicklas et al [57] reported a striking change in meal patterns over a 21-year period: They observed increases in the number of meals eaten away from home and at restaurants, decreases in home dinners, snacking and total eating episodes.

A recent review of current dietary trends and quality, including evidence of tracking of nutrient intake in children, as well as meal patterns, frequency and portion size information in US children aged two to eleven years has been compiled by the Ameican Dietetic Association (ADA) [16] and may also be relevant for older children and adolescents.

Hackett et al [58] studied eating habits of eleven and twelve-year-old children before and after the start of a healthy eating campaign in the United Kingdom (UK). They found favourable changes, but also that the "case for encouraging changes in the eating habits of children is compelling". This was confirmed by national and regional surveys of 11 to 14-year old UK children where the most popular items emerged were the least desirable foods: Confectionary (crisps, chocolates, sweets), biscuits and cakes, chips and sugar-flavoured fizzy drinks.[59] The dietary trends amongst Scottish school children in the 1990s suggest increased intakes of fruits and vegetables (but still below recommendations) and concomitant increases in high-fat and high-sugar foods, the latter particularly amongst boys and children from lower socio-economic groups.[13] A study of 158 German primary school children showed that they consumed 42% of energy from fat with about 50% as saturated fat.[60]

For South Africa national data for children are limited to the age group one to nine years.[61] Very few studies addressing children are included in a report on South African food consumption studies undertaken amongst different population groups between 1983 and 2000.[62] The THUSA Bana study focused on people in transition in the North West Province and included 1257 children, of which 868 in the age group ten to 13. Maize porridge, white sugar, brown bread, full cream milk and white bread were the most commonly consumed foods.[62] Over 40 years ago (1962) the nutritional status of six to eleven year-old white primary school children in Pretoria was surveyed in depth. At that stage percentage energy as fat, protein and carbohydrate were reported to be 35%, 12% and 53% respectively. It was concluded that the nutritional status was equivalent to that of an "affluent population group." [63]

2.1.2 Dietary fats in childhood nutrition

One of the US Dietary Guidelines 2000 for the general population states: "Choose a diet that is low in saturated fat and cholesterol and moderate in total fat." [12] The equivalent in the food-based dietary guidelines for South Africa is: "Eat fats sparingly." [13] The NCEP report of the Expert Panel on Blood Cholesterol Levels in Children and Adolescents [17] recommends the following intakes in the Step 1 diet:

- Total fat — Average of no more than 30% of total energy
- Saturated fatty acids — Less than 10% of total energy
- Polyunsaturated fatty acids — Up to 10% of total energy
- Monounsaturated fatty acids — Remaining total fat energy
- Cholesterol — Less than 300mg/d
- Carbohydrates — About 55% of total energy
- Protein — About 15-20% of total energy
- Energy — To promote normal growth and development and to reach or maintain desirable body weight

The American Heart Association Guidelines [15] reiterated the above as being population guidelines, which should also apply to children and adolescents.

Butte [64] reviewed the optimal fat intake for children against the background of their energy requirements. They state that the current recommendations of 30% of energy from dietary fat for children older than two years are sufficient for adequate growth. Lower intakes may be associated with micronutrient inadequacies. Higher intakes may lead to increased energy intakes and increases in body fat, but conflicting data are available.

In industrialised countries the conclusion thus seems to be that the primary prevention of CNCD should begin in childhood, [15, 16, 65] but there is no agreement on the most appropriate application, for example the US versus Canada regarding the best age from which to recommend these intakes.[66] In many developing countries too low fat intakes may be a greater concern,[67] leading several researchers [68, 69, 70] to point out the possible dangers of dietary fat restriction for children. Nevertheless, in a contra-point Lytle [71] defended a low-fat diet for healthy children and Van Horn [72] also reconfirmed the NCEP stand.

The Institute of Medicine in the most recent release of Dietary Reference Intake values seems to have accommodated both sides of the coin by formulating 'Acceptable Macronutrient Distribution Ranges' (AMDR), defined as 'a range of intakes for a particular energy source that is associated with reduced risk of chronic disease while providing adequate intakes of essential nutrients'. [10]

To address the nutritional excesses and deficiencies of South Africa, a land of contrasts, Vorster et al [73] have also suggested that more specific guidelines be adopted, for example instead of advising that less than 30% of energy should come from fat, ranges, for example between 25% and 30%, or to aim for 30% should be considered.

2.2 DIETARY ASSESSMENT OF CHILDREN

2.2.1 Overview

Against the backdrop of evidence that many of the risk factors for the development of CNCD, for example nutrient intake,[74] obesity [21] and hypercholesterolaemia,[75] track from childhood into adulthood, dietary assessment of children is important in nutrition monitoring, research, and in clinical and community-based interventions.

Stang [76] has compiled a practical overview of assessment of nutritional status in clinical practice, including dietary assessment, of adolescents. In the research context methods that can be used for dietary assessment range from very sophisticated individual-level investigations suitable for metabolic wards to 'bird's eye views' aimed at describing diet on group level. Bingham [77] has published a comprehensive review of the dietary assessment methods for use on individuals. Several overviews [30, 78, 79] focused their discussion on dietary assessment of children.

Apart from the general accuracy and precision issues of dietary evaluation,[80] studies in children pose additional challenges.[79, 81] Three recent reviews emphasised the importance of establishing reliability and validity of dietary assessments specifically in children.[25, 30, 81]

The *duplicate portion technique* has sometimes been described as giving very accurate information on the nutrient level. Isaksson [82] reviewed the principles involved. When total dietary fat and fatty acid intake measured by chemical analysis of duplicate diets were compared to nutritional database analysis of estimated dietary records, collected over the same three-day period, lack of agreement was found.[83]

Unobtrusive *observation* in assessment of children's dietary practices minimises self-report problems and has been considered as a ‘gold standard’ against which other measures of behaviour could be compared. The use thereof has been described by Baranowski and Simons-Morton.[79]

The four methods most commonly used for assessing diets of individuals are the food record, the 24-hour recall, the food frequency questionnaire (FFQ) and the diet history. The first two methods describe current intake and are meal-based, whereas the latter two describe past or usual diet. All of these have been used in assessment of diets of children.[25, 78] One of the major long-term studies involving children showed the feasibility of implementing a variety of dietary assessment methods among pre-adolescent children without relying primarily on parental reports, [84] but for younger children parents are usually included either as surrogates or in addition to the child report.

New approaches, for example using the computer, telephones and tape recorders to record children's food intake are being investigated.[85, 86] Diet analysis tools are increasingly available online.[87]

The FFQ and food record are discussed in more detail in later sections of this literature review.

2.2.2 Cognitive abilities of children affecting dietary assessment

Self-report of diet necessarily involves cognitive processes, although for many years limited research has focused on either adults' [88, 89] or children's [90] cognitions in regard to food. Recently various research groups have started paying attention to this aspect.

It can be assumed that children in the four major periods of cognitive development (that is sensorimotor [birth to two years], pre-operational [two to seven years], concrete operation [seven to eleven years] and formal operations [eleven years and beyond]) differ in the way they process food information. The latter age is about the age from which children have been shown to provide reasonably accurate dietary information.[25, 78, 81, 90]

In order to systematically analyse the mental activities involved in food recall, Baranowski and Domel [90] have proposed a model of a child's cognitive processing of food information (Figure 2.1). This model is the result of combining cognitive psychology with survey methodology in order to optimise the collection of valid food intake data. In the model the recall of foods and the number of portions of such foods is addressed, primarily on the short-term. The cognitive skills involved in recalling frequency of intake (for example event equalisation, estimation of frequency, averaging) are not explicitly covered, but can be inferred. The model analogises human cognition to methods by which a computer processes, categorises, stores, and retrieves information. Apart from the implications for developing new dietary assessment tools, this model provides a starting point for categorising errors that can be encountered with children's self-report of diet. As is evident from Figure 2.1, these errors can be related to attention, perception (or interpretation), organisation, retention, retrieval, and response (printed in italics in the Figure 2.1).

Noticing, that is paying *attention* to food eaten is a prerequisite for future recall. Inattentiveness may result in underreporting. The model also shows that paying attention to the request for dietary information is the critical starting point for valid recall. By increasing the interest in the task of remembering what was eaten, the quality of response can be improved.[90] Question comprehension is a critical cognitive stage in any dietary assessment involving recall. [91] In a FFQ it is essential that the participants understand the question and know how to report consumption frequency, portion sizes and compute average yearly use of seasonal items.[91]

The foods consumed must be *perceived* by the child to be the same thing that the researcher meant. In their study involving fifth to seventh graders, Koehler et al [92] listed food knowledge, preparation and vocabulary as instrument-related factors influencing the validity of a dietary assessment tool. In this respect, the use of pictures has been shown to reduce misunderstanding. Picture-to-picture matching appears to be superior to picture-to-word matching, and pictures appear to trigger memory, where words have not.[88]

Organisation, in the dietary recall context, refers to the grouping of foods in long-term memory. It may be that children classify foods differently to adults, for example by using functional criteria (that is meals versus snacks), nutritional or healthful criteria, or sweetness [90] instead of, for example, the basic food groups. Furthermore, different children may organise foods differently, use different reasons for categorisations, and this may be affected by developmental stage. Baranowski et al [90] and Koehler et al [92] reported that some children had difficulty understanding the wording of food categories on a food frequency form, for example when deciding in which category particular foods should be placed. Based on this rationale, some researchers (for example Kohlmeier [88]) have started rearranging their FFQ to a meal-based, rather than a list-based format.

Retention refers to memory and is related to time lapsed between the actual consumption and the request to recall the intake. In the case of children it was found that food memory decay varied by food group. However, underreporting appears to be more common than overreporting, even though this may, in part, be the result of researchers failing to differentiate between 'underreporting' (reporting one half of a banana when a whole banana was eaten) and 'failure to report' (reporting that no snack was eaten when in fact a banana was eaten).[90] Baxter and Thompson [93] found that the cognitive burden of recalling items eaten at school lunch as part of a 24-h recall was greater than that of recalling school lunch items as single meal. Thus the latter yielded more accurate information. Even under the best conditions (for example reporting within 90 minutes after the meal) children have difficulties reporting their intakes.[94] This is, however, not unequivocally accepted: In a longitudinal study Dwyer and Coleman [95] found that there was not necessarily a clear decline in accuracy of report over time when the same subjects were studied over four decades. Nevertheless, the distorting effect of current diet on recall of past food consumption was revealed.

The process of *retrieval* involves obtaining information out of long-term memory into short-term memory to form a response. In this stage interference can be a problem.[91] As the time interval is increased over which diet must be recalled, interference will also be increased. It is unlikely that food information is stored separately, but it is probably embedded in other events.[90] This is why the use of event prompts (for example a party, sports event) can improve a 24-h recall. Domel [29] explored how children remembered food intake and identified several memory retrieval-response categories, including visual imagery, usual practice, behaviour chaining, preference, food labels and so forth. The effect of different types of prompting on the accuracy of children's food recalls has consequently been investigated.[96] These researchers found that among first grade students, specific prompting in terms of preference, food category or visual cues resulted in more harm that good. Among fourth-grade learners prompting for food categories resulted in some improvement in accuracy. When children report eating standard portion sizes rather than the real amount eaten, this can result in overreporting of low intakes and underreporting of high intakes, the so-called flat slope syndrome.[90] Prompting children to report foods eaten over the previous 24 hours in reverse versus forward order improved omission and intrusion rates of fourth-graders' recalls, particularly for boys, but the overall error rate (omission plus intrusions) remained high.[97] Ensuring recognition of food items (either on a word-based list or on pictures), which is cognitively the core task in food frequency questionnaires,[89] aids the retrieval process.

Finally, *response* refers to the way in which children wish to present themselves to others. Social desirability plays a role here and was found to occur when children underreported candy consumption and over-reported vegetable consumption in a telephone recall compared to their parents' reports.[90] If an event is considered as embarrassing, sensitive in nature, threatening or divergent from the respondent's self-image, it is less likely to be reported.[91]

Thus, theoretically, memory and cognition are required for completing a FFQ, because participants must first recognise the food item. The consumption of each item must then be considered, the information over the reference period (for example one year) be integrated and finally the average frequency of food use computed. All of these processes are interlinked as indicated in Figure 2.1.

In practice Drewnowski [98] has, however, argued that "reality is beside the point: FFQ's reflect a long-established predisposition toward a mental image of a given food." He thus implies that the cognitive processes are overemphasised and that in actual practice food preferences or attitudes are measured by FFQ and not 'usual consumption'.

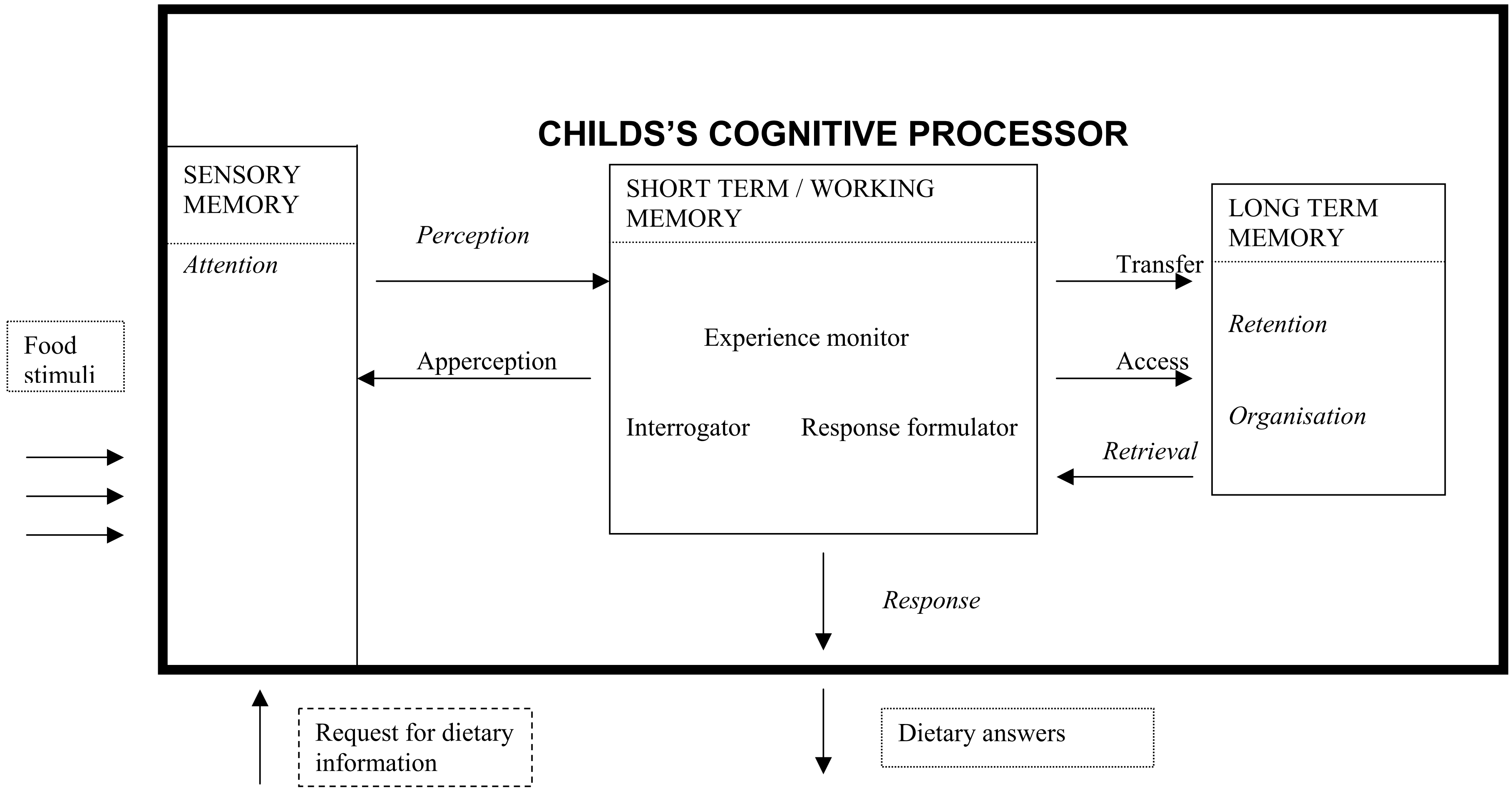

FIGURE 2.1: MODEL OF A CHILD'S COGNITIVE PROCESSING OF FOOD INFORMATION [90]

2.2.3 Integrating nutrition and dietary assessment into the school environment

Frank [55] and Story et al [99] identified multiple environments of children, all of which influence eating patterns and also the choice of method for collection of dietary information. Apart from the personal, the home, the media / entertainment environments, as well as fast food eateries and the food industry, they highlighted the school environment, where most children spend six to eight waking hours per day.

Schools have consequently been identified as ideal settings for promoting health and lifelong healthy eating amongst children for the following reasons: [65]

- Schools can reach almost all children
- Schools can provide opportunities to practice healthy eating
- Schools can teach children how to resist social pressure. Since eating is a socially learned behaviour, social (peer) pressures that discourage healthy eating can be directly addressed and positive peer pressure can be reinforced
- Skilled personnel are available. Teachers can receive nutrition knowledge and then use their instructional skills to reach the children
- Research suggests that school-based nutrition education programmes can improve eating behaviour of children [65]

The use of schools for achieving nutrition aims has repeatedly been documented in the international literature for primary, middle and secondary schools, in urban and rural settings, for CNCD risk reduction [100, 101] and for addressing and preventing obesity [102] and undernutrition, as stand-alone nutrition education projects and as integrated programmes where, for example, nutrition education, food provision and promotion of physical activity are jointly included in a bigger programme.[103] The Health Promoting Schools concept of the WHO is another example of integration.

Florencio [104] reviewed various approaches to school-based feeding programmes around the world. Those, which aim to tie-in with nutrition education range from a focus on protein-energy malnutrition, to micronutrient deficiencies and obesity. Sometimes the emphasis is on knowledge, but values, attitudes and skills have also been the objective. Strategies used include formal, class-room-based and also informal, extra-curricular activities, with varying emphasis on the involvement of the child, the existing teachers, parents and the community.

A review on using the school environment for promoting physical activity and healthy eating was published by Wechsler et al.[105] They concluded that "enough is known from theory, practice and research to suggest that school-based environmental strategies to promote physical activity and healthy eating among young people merit implementation and ongoing refinement". A practical, theory-based attempt to establish school nutrition advisory councils as an integral part of a school environment approach towards nutrition promotion has been proposed and described by Kubik et al.[106]

Many recent school-based nutrition programmes are placed within the framework of the US Center for Disease Control's recommendations for school health programmes promoting healthy eating. The seven recommendations are policy formulation, curriculum development for nutrition education, instruction for students, integration of food service and nutrition, training of school staff, family and community involvement, and programme evaluation.[65]

In South Africa the national Integrated Nutrition Programme (INP) reflects current policy in this regard. It includes the Primary School Nutrition Programme (PSNP), which aims to provide food supplementation, nutrition education and parasite control to the poorest of the poor, in an attempt to address short term hunger, the high prevalence of inadequate growth and micronutrient deficiencies (referring to iron and vitamin A), and to improve school attendance. Thus cooperation between nutrition (as part of the Department of Health) and the Department of Education is officially encouraged as part of intersectoral collaboration.[107]

Education in South Africa is in the process of moving towards outcome-based education (Curriculum 2005). This approach favours, amongst others, practical application, skills development and real life problem solving as desirable outcomes of education. The aim is to have integrated curricula.[108]

Nutrition education and assessment in the school context are, however, not without challenges: Teachers and administrators often view it as a loss of classroom instruction time.[55] This negative attitude can be reduced if the message and measurements are incorporated into classroom instruction.[55] Numerous possibilities exist for embedding or integrating nutrition into other subjects and various school subjects could be 'vehicles' for carrying nutrition content.[104]

Different teaching and learning strategies can be employed for doing this. Among the elements identified as contributing to the effectiveness of nutrition education and nutrition assessment for school-aged children are the following: [92, 104]

- The use of developmentally appropriate learning experiences and materials
- Activity-based teaching strategies
- Behaviourally focused approaches
- Educational strategies derived from appropriate theory and research
- Provision of adequate time, intensity and materials
- Involvement of parents/family

Depending on the specific circumstances, nutritional/dietary assessment in the school environment provides unique opportunities: These range from direct observation of eating behaviour, the use of surrogate respondents to checking accuracy of reports from menus, plate waste et cetera.[55] On the academic side, mathematics offers a unique, real-life opportunity to practice numerical and other cognitive skills. James and Adams [109] claim that nutrition science and mathematics form a natural partnership, since nutrition incorporates numerous mathematical procedures. In Figure 2.2 mathematics concepts and procedures (such as sorting, classifying, statistics, probability, estimation, rates and proportion) that support integration of nutrition are indicated emphasizing the commonalities between the disciplines.[109]

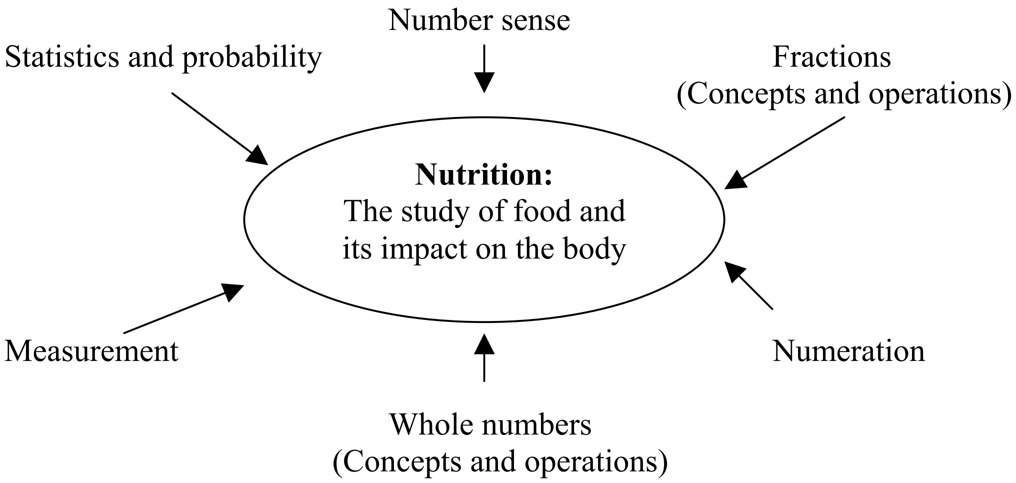

FIGURE 2.2 MATHEMATICS CONCEPTS AND PROCEDURES THAT SUPPORT INTEGRATION OF NUTRITION [109]

According to James and Adams [109] curriculum integration such as linking nutrition concepts to mathematics could have the following specific benefits:

- It encourages learners to use learning experiences to understand themselves and the world in which they live
- It engages the learner in searching for, obtaining and applying knowledge in a non-superficial way
- It provides learners with an opportunity to use their academic strengths to increase achievement
- It allows subject content and relevance to be viewed from various perspectives
- It supports natural, holistic learning
- A broader range of content can be presented in a meaningful way
- It may save time and money for teachers and school administrators

It is thus evident from international literature and from national health and education policy that integrating nutrition into the school environment is not only feasible, but also desirable.

2.2.4 The food frequency questionnaire (FFQ) as basic format of the test method

2.2.4.1 Description

A FFQ or checklist assesses dietary intake by determining how often a person consumes a limited number of foods. [25, 26] The original version was published in 1960 and was outlined as a 'short schedule for qualitative classification of dietary patterns'.[110] Kohlmeier and Bellach [111] have characterised the FFQ as being specifically designed to assess variance in the frequency of intake of particular foods, using a minimal number of closed questions. A respondent is presented with a list of foods or food groups (the item list), and then has to indicate how many times a day, week, month or year (s)he usually consumes these foods.[26] This requires that respondents add up frequencies of consumption across foods and consumption of individual foods across meals. [98, 111] Details about the FFQ are presented in most nutrition assessment textbooks (for example references [26, 27, 110]).

In some FFQ's a choice of portion sizes is not given; a ‘serving’ (that is standard portion size) is established from large-population data and assumed to be true. This is known as the simple or *non-quantitative or qualitative* food frequency questionnaire.[26] Hammond et al [112] used this approach for assessing dietary intake in five to eleven year old children. The *semi-quantitative* FFQ gives respondents an idea of portion size and requests that frequency of intake is provided in terms of this given amount.[26] An example is the widely researched Willet Questionnaire from

Harvard University.[113] Finally, the *quantitative* FFQ asks the respondent to describe the size of his or her usual serving as small, medium or large relative to a given standard. The Block Questionnaire from the National Cancer Institute is an example.[44]

2.2.4.2 Aims

The developmental aims of the quantitative (that is the most detailed) FFQ were: [44]

- Ranking of individuals by relative levels of nutrient intake and also estimation of absolute level of nutrient intake
- Representation of an individual's usual diet
- Relatively brief (for large scale use)
- Capable of assessing nutrients as well as food or food groups
- Assessment of a broad range of nutrients
- Assessment of a variety of demographic groups.

Since the original FFQ's were developed many variations have been designed and the FFQ is now the most popular dietary assessment tool in epidemiology. The comprehensiveness and detail contained in FFQ's vary greatly with regard to the item list, the nature, extent and time frame of the response for recording frequency of intake, and in terms of the reference portion size (from no quantification to sex and age-specific reference portion sizes). Furthermore, some FFQ's assess overall diet, whilst others are nutrient specific [26] and some assess dietary patterns. Consequently most FFQ's now stipulate own aims and usage criteria.

2.2.4.3 Strengths and limitations

The many variations of the FFQ may make generalisations regarding strengths and limitations invalid, but an outline of possibilities is given in Table 2.1.

TABLE 2.1: STRENGTHS AND LIMITATIONS OF FOOD FREQUENCY QUESTIONNAIRES (based on references [25, 26, 110, 114])

Strengths	Limitations
• Can be interviewer or self-administered (relative simplicity) • Trained interviewers not needed • Can be machine readable (if pre-coded) • Modest demand on respondents resulting in increased compliance (compared to other 'traditional' methods of dietary assessment) • Relatively inexpensive for large sample sizes • An indication of usual dietary intake may be obtained • Design can be based on large population data • Considered by some as the method of choice for research on diet-disease relationships (epidemiological studies) • Suitable for ranking or classification according to nutrient intake • Procedure does not alter habitual dietary habits Total diet or selected food or nutrients can be assessed • Response rates usually high (low respondent burden)	• Recall depends on memory • May not represent usual foods or portion sizes chosen by respondents because of incomplete or inappropriate listing of foods and errors in quantification • Intake data may be compromised when multiple foods are grouped within single listings • Not appropriate for determining absolute nutrient intakes like NHANES III • Development and validation is difficult, tedious and may be expensive • Limited data in terms of food descriptions • Period of recall imprecise • Respondent burden is governed by number and complexity of item list and quantification procedure • Recall of past diet may be biased by current diets • Heterogeneity of populations influences reliability of method (suitability questionable for segments of population consuming atypical diets or foods not on list) • FFQ with long list tend to overestimate and those with short lists tend to underestimate intake • No information on meal patterns throughout day • Considerable programming time and expertise required to convert food frequencies to nutrients

2.2.4.4 Development of food frequency questionnaires (FFQ's)

The basic design questions, which need to be addressed when developing a FFQ, include the following: [111, 114, 115, 116]

- Is information needed on foods, nutrients, dietary supplements or other food constituents, or specific behaviours?
- Which foods should be included?
- Which foods should be grouped together?
- Which consumption frequencies should be allowed?

- Is amount of consumption required? If yes, what should be set as a usual portion size
- How should individual foods be weighed in the development of the nutrient database?
- What should be the reference period?
- Is absolute or relative intake needed?
- What level of accuracy is required (individual versus group information for reproducibility and validity)?
- What are the constraints in terms of money, time, staff, and respondent characteristics?

Since cognitive psychologists became involved in FFQ construction, more design questions have evolved, resulting in studies investigating and designing the FFQ's and other dietary assessment methods from that perspective.[89, 117, 118]

Ideally, a FFQ's item list (that is the first three points in the above list) and the associated quantification (that is frequency of intake, portion size estimation and nutrient database) should be based on food intake data representative of the target population [44, 111, 112, 115, 119] in order to address content and measurement validity.

When Block et al [44] developed their FFQ (even though they then called it a self-administered ‘diet history’), they identified two fundamental questions that determine the performance of any FFQ:

"How accurate can the individual report on his frequency of consumption and his portion sizes?"
"How adequate is the food list itself, and its associated quantification?"

The latter is totally the responsibility of the investigator, and consequently instrument development, developmental evaluation and refinement are of critical importance to ensure the potential of precise measurement. Thus, during instrument development a methodological rationale aimed to address a limited, but essential question, namely: *"If respondents were able to respond accurately about their diet, could this food list and associated quantification adequately represent individual dietary intake, in spite of the diversity of dietary behaviour?"* should be provided.[44]

FFQ's should be adjusted for the population group for which they are intended,[116] since the context may well affect the participants' responses and validation characteristics.[120] The key questions to be considered before applying a FFQ designed for another population are the following: [111]

- Does the tool capture 80-90% of the interpersonal variance in consumption for the foods / nutrients under study? To do so, the questionnaire must be based on or compared with a recent survey of total diet assessed independently. This information must be conducted in the age, race, gender, ethnic, or religious group under study.[111]
- The categories (response options) of frequency of consumption also need to be examined for appropriateness for the group understanding.[111]
- Do the nutrient values attributed to each response apply in this group? Individuals consuming pumpkin pie will have incorrect carotene intakes if the nutrient data are based primarily on apple pie consumption. The nutrient value assigned to the questions should be a weighed average of up-to-date nutrient information on all consumed items that the question subsumes, because foods differ in some or all of their dietary constituents.[111, 121]
- Are there systematic biases in response between groups of people of interest? Is the accuracy of information captured, for example, in 50-year old African-American men similar to that from 50-year-old Caucasian men? Do 70-year-old men respond with a different degree of errors to the same questions than 20-year old men do? If so, applying the same FFQ to all people in studies spanning such different groups can result in an artefactual effect of diet.[111]
- Are the assumptions about portion size appropriate for the subjects under study in terms of their gender, age group and population? Are a significant number of participants in the study (for example elderly, vegetarians, children or members of a particular culture) for whom the assumptions on portion size are inappropriate? Use of a single set of portion sizes could then result in over- or underestimation of intakes.[111]
- If the FFQ is used to monitor changes in intake over time, how will change be accounted for? How will the introduction of new foods into the market, changes in price, changes in the use of specific food ingredients (for example oils used in manufacture of margarine) be handled?[111]
- Should a separate dietary assessment tool or biomarker be administered in a sub-sample of the population to calibrate the results from the FFQ and adjust for errors?[111]

2.2.4.4.1 Item list

Constructing the item list is the first step in the development of a FFQ and involves the decision which foods to include and which foods to group together. The aims of the FFQ will be the determining factor when the decision regarding which foods or food groups to include is made, and, as mentioned above, it should ideally be based on a recent survey of dietary habits in the target population.[111]

Since it is impossible to ask about all foods eaten, grouping is a core decision in the development of a FFQ. Block et al [44] used the following criteria for grouping or keeping foods separate:

- Conceptual similarities
- Respondents' ability to make the necessary distinctions
- Similarity in nutrient content per usual serving (not per 100g)
- Importance of a particular food to researchers' ability to correctly classify an individual with respect to nutrient intake
- Approximate number of persons at risk of such misclassification

The reductionism and summation, which are the logical consequence of the grouping, may lead to some loss of the real variance between subjects (in terms of individual foods consumed) and at the same time introducing between-subject variance (in terms of the food groupings).[122]

Little is known about the way in which children deal with individual foods or groupings, but Koehler et al [92] suggested that children more accurately recalled specific items rather than categories in their ‘Yesterday's Food Choices’ instrument.

2.2.4.4.2 Quantification

Quantification of a FFQ includes the portion or serving size attributions of each line item and also the assumed nutrient content of each. In addition, the measurement of frequency of intake should also be considered since all of these eventually determine the quantitative relevance of the result.

- ***Reference portion size and portion size estimation aids (PSEA)***

<u>Terminology</u>: A *portion* is the amount of food that a person reports as being the quantity usually consumed. There is no standard portion size and no single right or wrong portion size. By contrast, the terminology around *servings* originated in the USA and is a standard amount used to help give advice about how much to eat, or to identify how many kilojoules and nutrients are in a food. Serving sizes are specified for the Food Guide Pyramid, Nutrition Facts Label (on food packaging) and the Exchange Lists for Meal Planning.[123]

Criteria for serving sizes in the Food Guide Pyramid are: [123]

- Amount of food from a food group typically reported in surveys as consumed on one eating occasion
- Amount of food that provides a comparable amount of key nutrients from that group (for example amount of cheese that provides the same amount of calcium as 1 cup of milk)
- Amount of food recognised by most consumers (that is household measures) or that can be easily multiplied or divided to describe a quantity of food actually consumed
- Amount traditionally used in previous food guides to describe servings

Food label servings are defined by the US Food and Drug Administration as 'Reference amounts customarily consumed per eating occasion' and are stipulated in the 1990 'Nutrition Labeling and Education Act'. In South Africa the 'Foodstuffs, Cosmetics and Disinfectants Act (1972)' regulates labeling, but manufacturers define the serving size.

Exchange list servings are specific amounts of food that contain about the same amount of carbohydrate, protein and/or fat and energy as other foods on the same list. Serving sizes of different foods on the same list vary.[123]

Even in the USA the public appears to be confused by the terminology. In a survey amongst grade three to five year old children they were reported to use the words 'serving' or 'helping' to name a food portion.[85]

<u>Portion size in FFQ</u>: The quantitative and the semi-quantitative FFQ include an estimate of usual portion size consumed. In the consensus document for the use and development of FFQ [114] it has been stated that allowing subjects to estimate portion size in the completion of FFQ's is considered more advantageous than using average portion size. Suitable methods are the use of defined small / medium / large options or estimation of portion size using photographs.

It has been shown that people have different concepts of medium. Therefore portions need to be defined in quantitative (preferably weight) terms rather than qualitatively.[124] In Western societies a 'portion distortion' is common. Portion sizes both inside and outside the US home are increasing and super-sized portions are becoming the norm in fast food establishments [125, 126, 127] implying that medium portions as used on food labels and by health professionals need to be revisited.

Estimating portion sizes is a challenging task for respondents. Therefore the use of PSEA has been introduced,[128] including two-dimensional (2D) (drawings of real food, abstract shapes and household measures; photographs; graphics; package labels) and three-dimensional (3D) (household measures; real food samples; models; replicas; bean bags and cartons of varying sizes; common objects, for example domino, card deck, palm of hand, fist, tennis and golf balls) aids.[129, 130] A review on the validity of PSEA indicated that, overall, differences between the different types of PSEA were not statistically significant.[129]

Participants when judging portion sizes use many cognitive strategies. Furthermore, for different foods different PSEA may be most appropriate. Qualitative research suggests that respondents preferred aids that were similar in size and shape to actual portions consumed for liquids or amorphous foods, and for solid foods they preferred a ruler.[131] For snacks, bowls seemed to provide a means by which individuals could accurately estimate their consumption.[132] In cases where foods are irregularly shaped, for example muffins, estimation strategies that did not use PSEA were more effective and reference to 'large', 'extra large' or 'jumbo' was a preferable approach.[131] In a study involving an African population it was found that solid foods were better estimated than food with an amorphous appearance when using a food portion photograph book as PSEA.[133]

Training and exposure to PSEA seem to result in improved reported portion size accuracy but a great deal of estimation error remains.[130, 134, 135, 136]

Limited research included children. Goodwin et al [137] found that ten to twelve year olds were able to use portion size model booklets as an adjunct to food recording, but the children reported that they would be reluctant to use it when their peers were around. Frobisher and Maxwell concluded that for subjects 16 years and younger alternatives to the food atlas and descriptions would be more appropriate (for example standard food portions) as they observed high error rates.[138] Assignment of a standard serving size was also recommended for dietary assessment of young children where caregivers acted as data source.[139]

- ***Frequency of intake***

In most cases the main determinant of variation in measuring dietary intakes is frequency of consumption of the food items in the list,[140] but in certain contexts (for example rural areas of Korea) between-persons variation of food such as cooked rice might be determined by portion size rather than by food frequency.[141] The frequency of intake on a FFQ can be detailed, that is requesting that a participant indicates the number of times each line item is consumed in a specific time period (per meal, day, week et cetera) or the response options can be categorised, meaning that a participant has to check intake in limited, preset categories typically in a fixed time frame (for example per week).

The cognitive tasks involved in providing a correct response on a categorised quantitative food frequency questionnaire starts off with recognizing the items on the list.[89] This is followed with at least the following (based on reference [88]):

- Recalling own intake
- Counting number of times per week foods are eaten
- Summation of counts of foods belonging to a food category
- Conversion of summated intake to a frequency of intake format used in the data collection tool (that is coding and recording the answer in the required questionnaire format).

Smith reviewed the cognitive psychological aspects of relevance to reporting FFQ's, including the credibilty of frequency judgments, the difference between rates versus counts and the cognitive implications of item grouping. It was concluded that many of the assumptions underlying the use of the FFQ should be subjected to scrutiny.[89]

- ***Nutrient content of items***

The nutrient database consists of the foods that have been determined to contribute the most to the variance in intake of the nutrient of interest in the population under study.[111]

Nutrient calculations for FFQ's are based on specially constructed nutrient databases that apply weighted averages of the proportions of intakes of all foods covered by the questions. Determination of average nutrient values requires up-to-date information on the relative consumption of the individual items as a proportion of the group under question in the population of interest. For example, if middle-aged African-American men eat more pork than Caucasian men of the same age, the values for thiamin for the response of daily consumption of beef, pork, or lamb will be based on a different database that weighs pork more heavily that the amount for a daily response for Caucasians. These assumptions about relative consumption need to be checked regularly and updated if necessary.[111]

Some of the methods used to assign nutrient values to food groups in FFQ's have been evaluated. In general, mean-based methods appear to be superior to median-based methods, but among the mean-based methods no one variation was consistently better.[121]

2.2.4.5 Food frequency questionnaires (FFQ) for children

When the FFQ is used for school-aged children it should be adapted. The following should be addressed: [142]

- The food list - If composites are used it is important to make sure that the children know which items are included
- Time interval - Time intervals need to be fixed by meaningful start and end points, and may need to be abbreviated
- Response set - Children may respond affirmatively to authoritatively phrased questions or they may adopt a response set when they are unsure, have no opinion, or are disinterested
- Context of questioning - Language needs to be consistent with the child's understanding
- Structuring of the questionnaire - Begin with easy questions on topics of interest with threatening or difficult questions last

McPherson [25] reviewed a total of 21 studies where the FFQ was validated in school-aged children. Of these twelve focused on specific food groups or nutrients and nine assessed the general diet. Parents served as proxies for children in six studies and assisted children in five studies. In one study [48] the accuracy of children's versus parents' responses was measured. The FFQ provided a better appraisal of eleven to twelve-year old children's intake when administered to parents rather than the children: overestimation of energy was more severe for children than for parents. No consistent patterns emerged for either age or gender.

Bellu et al [119] validated a 116 item FFQ completed jointly by Italian parents and their nine to twelve-year-old children by comparison to a seven-day dietary record kept by the parents. The overall validity of individual estimates was “fair for some but unsatisfactory for many nutrients”. The FFQ tended to overestimate intakes but calculating nutrient densities reduced the difference. When the FFQ was compared to 24-hour recalls obtained from the mothers the performance improved on average.[143] This highlights the challenges regarding the choice of a reference method when no 'golden standard' is available.

In a Belgian study the test-retest reproducibility and relative validity of a 15-item FFQ was assessed in three separate studies involving school-aged children aged eleven to 18 years.[144] The FFQ was found to be sufficiently reliable and valid to be useful for ranking subjects and could thus be used for multivariate or correlation analysis in epidemiological studies, but not for estimating prevalences.

In a small (n=20) sample of Swiss adolescents aged nine to 19 years Cavadini et al [145] found that a semi-quantitative FFQ administered at school during a one-hour session "correctly described food consumption" when compared to a modified diet history. Energy and macronutrient intakes estimated from a FFQ (Block98) and a three-day diet record in young girls (four to nine years of age), primarily by their parents, disagreed. It was found that the FFQ overestimated intakes.[146]

2.2.4.6 Food frequency questionnaires (FFQ) for fat intake

Dennis et al examined the quality of FFQ's aimed to assess the relationship between fat intake and prostrate cancer. A total of 39 studies met their inclusion criteria. From these studies they compiled a scoring method for evaluating FFQ's in general. They concluded that, whilst the FFQ has often been used to measure fat intake, methodological flaws in instrument development might partly be responsible for the inconclusive results regarding the relationship between dietary fat intake and prostate cancer.[147]

Studies in which the validity of the FFQ for specifically measuring fat intake was studied include the following:

- A 104-item FFQ aimed to assess fat and cholesterol intakes was found to correlate with biomarkers (linoleic acid in erythrocytes and adipose tissue) and a diet history in 191 adults.[148] Additional evidence that a FFQ can provide informative measurements of dietary fat was published by Willet et al,[149] who used plasma fasting triglyceride levels as an "alloyed gold" standard.
- Validity and reliability of a self-administered FFQ, designed to be sensitive to low-fat, regional and ethnic dietary patterns, were adversely affected when it was administered to minority or poorly educated populations.[150]
- A semi-quantitative FFQ did not provide reliable estimates of actual absolute or percentage fats or cholesterol in subjects consuming diets of known composition. [151]

- Incorporating fat-modified foods into the Block FFQ's item list improved the classification of fat intake when four two-day food records over a one-year period acted as reference method.[152]
- A FFQ specifically adapted to measure fatty acid intake was validated against seven-day weighed records and found to be a reliable estimate of dietary intake of individual fatty acids.[153]

2.2.5 The food record

2.2.5.1 Description

In this method of dietary assessment the participant records, at the time of consumption, the identity and amounts of all foods and beverages consumed for a period of time. Food and beverage intake can be quantified by estimating portion sizes, using household measures (that is *'estimated food records'*), or weighing the food or drinks on scales (that is *'weighed food records'*). Certain items, such as eggs, apples or cans of cool drink may be recorded as units or simply counted.

The food record does not depend on memory because the participant ideally records intakes at the time of consumption. This is in contrast to the FFQ where the task is to recognise food on a list and then remember and calculate usual frequency of intake. Thus, the error structure of the food record differs from the FFQ. The food record can provide detailed food intake information about eating habits (for example when, where and with whom food was eaten). Multiple-day data is more representative of usual intakes and non-consecutive, random days (including weekends) covering different seasons are necessary to arrive at useful estimates of usual intake.[26, 110, 154]

2.2.5.2 Strengths and limitations

The strengths and limitations of food records are summarised in Table 2.2.

TABLE 2.2: STRENGTHS AND LIMITATIONS OF FOOD RECORDS (based on references [25, 26, 110])

Strengths	Limitations
• Do not depend on memory • Defined record time • Intake can be quantified • Theoretically more accurate • Can provide detailed intake data • Can provide data about eating habits • Multiple-day data more representative of usual intake • Reasonably valid up to five days • Training can be group-administered • Procedure (technical instructions) can be automated	• Require high degree of cooperation • Recorder must be literate • Response burden can result in low response rates • Take more time to obtain data • Act of recording may alter diet • Data collection and analysis are labour intensive and expensive • Food eaten away from home less accurately recorded

A food record requires that that children can write names of foods legibly, recognise and describe quantities in either fractions or whole units, decipher food label information, and retain the record in their possession for completion of all entries during a day.[155] Children ages ten to twelve years are reliable respondents; adolescents are capable but often less interested in participating than younger children; Standardised procedures for completing the record reduce respondent error and food illustrations or flow sheets assist the recording technique.[155]

Wold et al were able to show that instructional flowcharts helped participants keep three-day records in terms of accurately describing food intake. Such records tended to be complete and specific.[156]

2.2.5.3 The food record as reference method

Some authors have referred to the food record as the 'practical golden standard' in dietary assessment.[110, 157] From the above-mentioned advantages it is evident why, over the years, it has extensively been used as the reference method in comparative validation studies where the even more costly methods such as duplicate portions,[125] direct observations,[158] doubly labeled water,[159] or other biomarkers [160, 161] were impractical or too expensive.

A key issue in validation studies is the independence of errors between methods.[154, 160, 161] From the summary of advantages and limitations of the FFQ and the food record, it is evident that the error structure of the food record is reasonably different from the error structure to be expected from the FFQ, making it a suitable reference method.

In the case of individuals, the food record measures current or actual intake.[162] Thus, if usual (that is habitual) intake of individuals is to be measured, multiple, non-consecutive days covering the reference period (that is all natural and trade / cultural seasons) are recommended, [26, 163, 164] since average intake reported in a long series of food records has been considered an operational definition of 'usual diet'.[157]

An unresolved question regards the number of days that should be recorded in order to assess usual intake.[165, 166] Carroll et al [157] investigated whether it was better to obtain many food records from a moderate number of subjects or a small number of food records from a larger number of subjects when the food record is used as 'gold standard' in a validation study. They concluded that neither strategy is always preferable: The aim of the validation study is of prime importance. For estimating correlations or slopes (as in the present study) within-person variance of the food record and distribution of true usual intake seem to be deciding factors.

The number of days of food intake records required to estimate individual and group nutrient intakes with defined confidence ($P<0.05$) have been published and those of relevance to this study are indicated in Table 2.3. The group size on which Basiotis et al [167] based their findings was relatively small (n=29) and they conclude that fewer days would be needed for larger groups. Overall, energy intake seems to require the least number of days to classify 80% of the population into tertiles of nutritional intake with a 95% confidence interval, [166] and fat seems to have relatively low intra-person, short-term variability and tends to remain stable in healthy free-living people.[168]

TABLE 2.3: AVERAGE NUMBER OF DAYS REQUIRED TO ESTIMATE TRUE INTAKE OF INDIVIDUALS AND GROUPS OF INDIVIDUALS (selected information from reference [167])

Component	Individuals		Groups of individuals	
	Males	Females	Males	Females
Food energy	27	35	3	3
Fat	57	71	6	6
Saturated fat	71	87	8	7
Cholesterol	139	300	13	15

Beaton [122] was able to show that for many nutrients, increasing the number of days of data beyond three does not materially contract the distribution, but it remains important to remember that this recommendation refers to a large sample of subjects and a true distribution of usual intakes. Relatively fewer days of recording appear to be necessary to describe the intakes of younger persons and people who are characterised by diets with limited diversity.[26, 166] However, if the aim of a study is to capture food variety in school children, then a two-week period appeared to be necessary for US fourth and fifth graders.[169]

Another issue in food recording is the decision which days to include. In order to represent all days of the week proportionally, the classical approach was the seven-day-food record.[110, 170] Since longer recording periods may reduce accuracy and cooperation, [171] but weekend days may be different from weekdays, ensuring that weekdays and weekend days are included is typically recommended.[26]

It appears that the variation in diet on different days of the week could be population-specific. Jula and co-workers studied the influence of days of the week on reported food, macronutrient and alcohol intakes in Finnish adults. They found relatively little variation on weekdays (excluding Fridays) but clear differences on Fridays, Saturdays and Sundays. They thus recommended five days of recording including any two days from Monday to Thursday, plus Friday, Saturday and Sunday.[172]

A South African study from the mid 1960's compared Indian, black and coloured children in Pretoria. A rank order of correlation coefficients between daily intakes of twelve nutrients on each day of the week and the average weekly intake revealed the following: Weekend days did not necessarily give a poor representation of a week's average intake, but certain days of the week may give a better indication of average daily intake for the week than others. For Indian children the highest correlation coefficient was found for a Thursday. Weighing food intakes on a Friday gave a more representative picture of the average daily intake of black children for the week than weighing on any other day, whilst for coloured children there was no difference from day to day.[173]

Research has indicated that dietary reporting decreases during the recording period. Therefore it has been recommended that in multi-day records the starting days should be evenly distributed across the days of the week to counter recording fatigue, boredom and training effects.

Otherwise the introduction of a systematic error with repeated measurements could outweigh any advantages of repeated records.[81]

From the above it is evident that many factors influence the decision about how many and which days to record, including the purpose of the study, the sex and age and typical diet of the group to be studied, and the nutrients of interest.[26] In general, it is recommended to include measures of quality control when collecting dietary data.[81]

2.2.5.4 The food record for children

A review on dietary assessment methods among school-aged children included six studies where the validity of the food record was studied.[25] The ages of the children varied from eight years (where adult assistance was required) to 19 years. In general, the authors reported that food records underestimated energy intake when compared to doubly labeled water. Few studies evaluated children's ability to complete the record on their own or to record an entire day.

The following additional studies investigated the validity of the food record in children:

Jenner et al [48] compared various methods of dietary assessment amongst Australian school children and found that two or three-day, carefully administered and thoroughly checked, food records were reasonably valid means of assessing usual intakes in eleven and twelve year olds. They reported consistently higher correlations and smaller differences between food records and their reference method (14 food records collected over several months), compared to any other (food frequency type) reference method tested. Their test and reference method were, however, both food records and 'auto-correlation' between methods that have the same error structure may have played a role, particularly because the two- or three-day record series was included in the 14 day series.

In the NHLBI Growth and Health Study nine- and ten-year old girls were assigned to one of three dietary assessment methods (24-hour recall, three-day record and five-day FFQ). At the same time unobtrusive observers recorded types and amounts of food eaten during lunch. It was found that the nature of errors in food reporting and quantification varied with the assessment method. The three-day record was reported to have a comparative advantage over the others in this age group.[174]

2.3 NUTRITIONAL AND DIETARY SCREENING

2.3.1 Definition and characteristics

In general, the WHO views screening as the use of presumptive methods aimed at detecting unrecognised health risk or asymptomatic disease in order to permit timely intervention (Braveman & Tarimo, 1994:6). This definition is very similar to that quoted by Rush: [175] It is "*the examination of asymptomatic people to classify them as likely, or unlikely, to have the disease that is the object of screening. People who appear likely to have the disease are investigated further to arrive at a final diagnosis*".

The ADA defined nutrition screening as "the process of identifying characteristics known to be associated with nutrition problems", [176] and the Joint Commission for the Accreditation of Health Care Organizations of the US and ESPEN of Europe have formally acknowledged its importance in patient-oriented, high standard nutrition care.[26, 36]

Screening is often the first step of nutritional assessment (level 1 nutrition care) provided by qualified health care professionals in hospitals, clinics, private practice and community settings with the aim to (i) identify children at risk, (ii) refer children at risk to a registered dietitian for in-depth nutritional assessment (level 2 care), and (iii) provide anticipatory dietary guidance and educational materials to families regarding prevention of nutritional problems.[177]

Splett and colleagues [178] consider screening as a trigger event, which initiates the nutrition care process. Thus it forms the access point for referral to nutrition care. The purpose of nutrition screening is to predict the probability of a better or worse outcome due to nutrition factors.[36] It can happen either during a general health or disease-focused screening, recognizing that a nutrition-related need can be identified as a potential or early risk factor or as a complicating or underlying factor related to an existing medical condition or disease. In disease-focused screening, nutrition problems can be a cause of a result of the disease, and they can have a physiologic or behavioural aetiology.

According to the ADA the nutrition screening process is characterised by the following: [176]

- it may be performed in any setting (given the opportunities and constraints of the intended venue, for example hospital-based versus tools for the home dwelling) [175]
- it aids in achieving early intervention goals

- it includes the collection of applicable data on risk factors and the interpretation of data for treatment or intervention
- it determines the need for a complete nutritional evaluation
- is cost-effective.

Rush [175] and Kondrup et al [36] specify that the usefulness of screening tools and programs should be evaluated by assessing thier predictive and content validity, reliability, practicality and link to action protocols. Furthermore there should be significantly greater benefit from earlier intervention than from what would result from intervention at the time the subject seeks help because of symptoms. Finally, screening should be shown to be preferable to other strategies, such as universal application of an intervention.

Nutrition screening can include anthropometric, biochemical, clinical and / or dietary data [26] and, in the case of young children, behavioural and skill development have been added.[177] Since there is no fixed boundary between screening and diagnostic testing [175] any nutritional status indicator can potentially be used for screening. The limitations are set by the complexity, cost and utility of the screening tool or protocol, by the prevalence of the problem being assessed and the potential benefit from the intervention. Thus dietary screening is a form of brief nutritional assessment focussing on food intake.

Interest in dietary screening is based on the observation that a large fraction of the variability of nutrient intake can be explained by a small number of foods.[179] Simple tools designed for dietary assessment (for example HEA1, HEA2, HEA3, DINE, Nurse Questionnaire) were shown to perform as well as much more complicated and time-consuming tools (for example 24-hour recalls; checklists) and their performance may even be comparable to the seven-day record. [120] There have been reports where a FFQ using seven broad categories correlated more highly with reference values than a FFQ using 31 individual fruit and vegetable items.[180] For children, methods of dietary assessment that are perceived as being less burdensome and time-consuming may improve compliance.[81]

Some short assessments are part of larger surveillance programmes, for example the ‘Youth Risk Behavior Surveillance System’ for American adolescents, which includes seven questions about the previous day's food choices.[181]

2.3.2 Aims

In general, there are different types of screening, each with specific aims: [45]

- *Mass screening* involves the screening of a whole population
- *Multiple* or *multiphase screening* involves the use of a variety of screening tests on the same occasion
- *Targeted screening* of groups with specific exposures
- *Case finding* or *opportunistic screening* is restricted to patients who consult a health practitioner for some other purpose

Wilkin et al [182] divide the purpose and use of health-related measures into three broad categories, namely discrimination, prediction and evaluation. According to these researchers there may be more than one purpose for a tool. In the case of *discrimination*, the purpose of the measure is to classify individuals or groups based on some health-related dimension, for example as a means for identifying areas of need, or to help target those whose needs are greatest. The purpose of *prediction* is to identify groups or individuals who have or will develop some target condition or outcome. It is thus aimed at predicting future need at an early stage in order to save time or costs, or to be predictive of a more detailed assessment. *Evaluation* is intended to measure or monitor the magnitude of longitudinal change in individuals or groups on the dimension of interest, for example by focusing on changes over time attributable to an intervention.

Four fundamental measurement axioms have been proposed for any tool used for health measurement. De Vos [183] summarised these as follows:

- If an instrument must have any utility in practice it must be valid and reliable
- For maximum utility an instrument must be brief, easy to administer, easy to understand, score and interpret
- There are only two ways to determine whether a client / patient has a problem: watch him or ask him. Thus direct observation and client report are the methods in which information can be obtained
- There are only four ways of measuring client / patient problem: in terms of its switch, frequency, magnitude or duration. Switch refers to presence albeit absence of the problem. The frequency is obtained by assessing how often the problem is encountered. Magnitude or intensity characterises the degree to which the problem is present, and duration specifies the length of time the problem is continually present.

Each of the above axioms is seen to have general and specific implications for the development of a dietary screening tool as a health measure.

According to Keller et al [184] and Jones [185] the criteria for the tool must be stipulated by specifying what the tool is intended to achieve in which population. In this context it is important to clarify whether the measure focuses on individuals or groups.[182]

2.3.3 Examples of screeners

Some screening tools are intended to pick up general nutritional risk of adults (for example references [186, 187]). A computerised diet questionnaire for the use in health education aimed at giving rapid feedback to the general public was validated against 16 days of weighed diet records. At least 65% of subjects were classified to within one quintile of the classification of the record for most of the nutrients assessed.[188]

Another group of screeners may be aimed at specific target groups and contexts. Examples are the numerous tools that have been developed to identify nutritional risk of the hospitalised patient on admission for example the 'Derby Nutritional Score' [189], the 'Veterans Affairs Nutrition Status Classification' [190], nutritional scores such as the 'Prognostic Nutritional Index' (PNI), the 'Nutritional Risk Index', the 'Subjective Global Assessment' (SGA), the 'Mini Nutritional Assessment' (MNA), the 'Registered Nurses Nutrition Risk Classification' and others.[191, 192, 193]

Several dietery screeners that measure fruit and vegetable consumption have been published.[194, 195, 196, 197]

The 'Family Eating and Activity Habits Questionnaire' is an instrument that identifies factors in a child's family environment that facilitate obesity.[198]

2.3.3.1 Fat screeners

In two publications [199, 200] an overview of short or qualitative questionnaires assessing fat intake has been provided.

A number of fat screeners used or included elements of a FFQ. These include:

The original MEDFICTS instrument as recommended by the NCEP is a simple approach for rapidly assessing a person's adherence to the Step 1 and 2 diets,[26] making it an efficient tool in cardiovascular screening, clinical practice, or research aimed at detecting individuals consuming diets above or below the cut points set out in the ‘Heart-Healthy’ (Step 1) and ‘Therapeutic Lifestyle Change’ (TLC) (Step 2) diets. It was originally developed for Adult Treatment Progam (ATP) II and has again been included in ATP III.

In a pilot test validation the MEDFICTS score was significantly correlated with percentage energy from fat ($r=0.8$, $P=0.0002$), percentage energy from saturated fatty acids ($r=0.8$, $P=0.0003$), and dietary cholesterol ($r=0.5$, $p<0.05$).[201] Similar results were obtained when MEDFICTS was self-administered or nutritionist-administered, and then compared to recent three-day food records.[202] More recently Kris-Etherton et al [203] confirmed the validity in follow-up studies. The MEDFICTS dietary assessment tool has also been adjusted to accommodate cultural differences.[204] Taylor et al [205] validated the tool amongst adult army recruits.

Block et al [206] developed a 13-item FFQ type dietary screener for high fat intake. They found that among 101 females aged 45years and older, that the tool performed nearly as well as a four-day diet record in correctly identifying those above and below the group midpoint in PFE. Caan et al [49] modified this tool and evaluated its sensitivity, specificity and predictive value. They found variations in these indicators of validity leading them to conclude that it could not be used as a single assessment method.

A qualitative fat index was validated with a three-day food record as reference method. The index was based on four questions, which reflect the most important sources of fat in the Finnish diet. This was supplemented with a short FFQ consisting of 21 items. The latter proved to be accurate at group level and the former for measuring quality of fat.[207]

The ‘Fat List’, a short FFQ, was compared to the seven-day food record. For Dutch adolescents the correlation between the two methods for total and saturated fat intake in grams was 0.6. For percentages energy from fat the correlations were low.[199]

Murphy and colleagues developed a food behaviour checklist for use in low-income (EFNEP) groups and evaluated its criterion validity using biomarker and convergent validity with multiple 24-hour recalls. Overall, the fat and cholesterol-related items performed poorly, their internal consistency was low and correlations with PFE were weak.[208]

A behavioural approach or combinations of methods formed the basis of a number of other dietary fat screeners:

The 'Food Habits Questionnaire' of Kristal et al [209] is a behavioural approach to assessment. The 18-item scale had high reproducubility and internal consistency and correlated well with PFE in middle-aged females. Birkett and Boulet [210], however, found poor performance amongst male labourers in terms of reliability (Cronbach's alpha and item total correlations) as well as validity measured with partial correlations.

In the Family Heart Study, a coronary heart disease prevention project, an "inexpensive, reliable and valid" instrument for rapid assessment of eating habits and diet composition was used, consisting of 32 items. The researchers measured validity by comparison with 24-hour dietary recall and by comparing changes in diet with changes in plasma cholesterol levels in a five-year period.[211]

'Rate your Plate' is a brief eating pattern assessment and educational tool used for cholesterol screening and education programmes. The authors stipulate that it is neither a measurement of usual, long-term, nor of quantitative intakes.[212]

The 'Dietary Risk Assessment' (DRA), originally developed by Ammerman et al [213] to identify dietary behaviours associated with cardiovascular disease, was compared with multiple 24-hour recalls and a seven-day recall. The correlations were moderate, but it was recommended as a primary care screening instrument for higher fat intakes.[214] Dietary behaviours related to total fat and saturated fat intake have been identified by Capp et al.[215] The results were expected to have implications for designing brief fat assessment instruments.

A saturated fat / cholesterol avoidance scale consisting of six component items was developed and its internal consistency and criterion validity (relative to scores on the Keys equation and self-report of diet by means of 24-hour recall, FFQ and fat behaviour) were determined. It was recommended as a useful tool in epidemiological research on cardiovascular risk factors. [216]

A twelve-item questionnnaire ('Fat Habits Score') has been developed to evaluate group changes in fat intake. The score was compared in children and adults with estimates of saturated and total fat intake (percentage of total energy) from a FFQ. Both questionnaires were re-administered six months later and it was found that the simple score was able to detect changes in fat consumption (Kinley, 1991).

The 'Fat Intake Scale' (FIS) consists of twelve items related to dietary fat, saturated fat and cholesterol. It was compared to food records and the score was found to have acceptable reliablity and validity. The Keys score and the RISCC (ratio of ingested saturated fat and cholesterol to calories) score were additional diet scores with which the FIS correlated.[217]

The 'Food Behaviour List' has also been developed by Kristal et al.[218] It is a simplification of the 24-hour recall that consists of 19 yes/no questions about foods consumed the previous day. Its agreement with a professionally administered 24-hour recall was tested. Preliminary evidence suggested that it was a valid measure of lower fat-higher fat intake.

The 'Diet Quality Index Revised' [219] reflects adherence to current dietary guidance by populations. Three of the ten components of the tool relate to fat intake (that is total fat =<30% of energy; saturated fat =<10% of energy; dietary cholesterol <300mg).

From the above it is evident that numerous screeners for dietary fat intake have been published. Very little attention has been paid to dietary fat screeners for children.

2.4 RELIABILITY AND VALIDITY IN DIETARY ASSESSMENT AND SCREENING

One of the requirements of dietary assessment and screening tools is that they should be reliable and valid. In the quantification of reliability and validity a distinction should be made between variability and error.

2.4.1 Variability and error in dietary assessment

2.4.1.1 True variability

Dietary intake is characterised by a 'true variability', which includes both intra- (within) and inter- (between) subject variation. Since this variation characterises true usual intake, no attempt should be made during the measurement of diet to minimise this variability.[220] Instead, researchers are encouraged to design their projects in such a manner that these two sources of variability can be separated and estimated statistically. In this way the magnitude of the effect of intra- and inter-subject variation can be taken into account during the interpretation of the data.[220]

Variability within the individual may (i) occur from day to day (that is diurnal variation, for example the day of the week effect in dietary intake), [220] (ii) follow a consumption curve (for example natural and commercial seasonality) [163, 221] or (iii) progress with normal growth and development. Gibson [220] adds to this a training effect where a subject alters intakes in reaction against repeated interviews.

This intra-subject variation is particularly important if data on usual intakes are to be correlated with other parameters (for example biochemical or clinical findings), since large intra-individual variation in intake will tend to reduce the absolute value of the correlation. The resulting attenuation of the correlation coefficient could, for instance, be a reason for the apparent lack of a significant relationship between dietary fat intake and serum cholesterol levels in individuals.[220]

Variations within populations (inter-individual variability) can be considered the cumulative variability of individuals, and, generally speaking, knowledge about variability in populations makes it possible to define ranges of 'normality'. Environmental (for example geographic) and genetic influences play a part in this regard. In addition, age and gender are sources of inter-individual variation that need to be considered in measuring diet.[220]

If inter-subject variation is large relative to intra-subject variation, subjects can be readily distinguished so that usual nutrient intakes of individuals can be characterised. However, for most nutrients, inter-subject variation is smaller than intra-subject variation, and consequently, mean intakes of groups can be measured more precisely than individual consumption.[220] The ratio of intra- to inter-individual variance is nutrient-specific (for example when based on 24 days of records, the within- to between person variance ratios ranged from 1.4 for saturated fats to 4.6 for vitamin A).[111] Similarly, the precision estimates from one 24-hour recall in estimating energy intake for a typical male would be ±51%, whereas it is ±293% for vitamin A.[26] Gender, age, ethnic group, and country are also known to affect the ratio of intra- to inter-individual variance.[26, 111]

It follows that knowledge of the true variability of the attribute of interest (in this case habitual fat intake) is important in order to ensure measurement of the true picture of usual consumption in a particular individual or population.

2.4.1.2 Error

Whilst true variability should be reflected by dietary assessment, measurement errors (which can be due to poor calibration of the instrument, inherent lack of precision of the instrument, or mistakes in the collection, reporting and recording of information by the subject or researcher/dietitian) [222] should be controlled and minimised.[220] Errors associated with the compilation of nutrient data and the nutrient analysis of food items are another source of error, which should be kept in mind.[27]

Two types of measurement error can be distinguished in the measurement of diet: systematic and/or random. [91, 166, 220, 222]

Systematic error occurs when there is a tendency to produce results that differ in a systematic manner from the true values, that is a systematic under- or overestimation in an individual or groups of individuals.[122] It is formally defined as "any process at any stage of inference, which tends to produce results or conclusions that differ systematically from the truth." [223] A study with a small systematic error has a high accuracy, independent of by sample size.[45] Since systematic errors reflect bias, the control (and ideally elimination) thereof should be addressed during the testing and validation of a technique, because they cannot be removed by subsequent statistical analysis.[220]

Over 30 types of specific types of systematic errors have been identified in epidemiology. The two most important examples are, according to Beaglehole et al [45], selection and measurement (classification) bias. Selection bias occurs when there is a systematic difference between the characteristics of the people selected for a study and the characteristics of those who are not. Beaglehole et al indicate that measurement bias occurs when the individual measurements or classifications of disease are inaccurate, for example when different laboratories produce different results on the same specimen. Confounding (which arises when the non-random distribution of risk-factors in the source population is also present in the study population) is sometimes added to the systematic errors even though it is not the result from a systematic error in research design.[45]

In nutrition epidemiology selected examples of bias that may apply to assessment of exposure are summarised in Table 2.4. [223] The examples in the Table are errors due to the respondent or the interviewer.[27, 91]

TABLE 2.4: BIAS THAT MAY OCCUR IN NUTRITIONAL ASSESSMENT (based on reference [223])

Type	Description	Comments or examples
Insensitive-measure bias	When outcome measures are incapable of detecting clinically significant associations	May reflect difficulties in accurate recall, portion estimation, and generalisation to 'usual diet'; Relevant to recall methods and particularly FFQ; Unintentional [88]
Underlying-cause bias (recall bias)	Cases may ruminate about possible causes for their illness and thus exhibit different recall to previous exposure than controls	In case-control studies where diet is assessed retrospectively
Unacceptability bias	Measurements which embarrass or invade privacy may be systematically refused or evaded	Obese subjects may be prone to this type of bias
Obsequiousness bias	Subjects may systematically alter responses in the direction they perceive desired by the investigator	In face-to face interview situations; The risk of intentional wrong answers increases if the subject believes that a quality scale is involved, for example the 'desirable' responses of FFQ maybe perceived to be either on the left or the right side of the form [88]
Expectation bias	Observers may systematically err in measuring and recording observations so that they concur with previous expectations	In interviews where unusual diet is reported; Following an intervention, participants bias their reports to appear in compliance with the intervention goals [224]
Exposure-suspicion bias	A knowledge of the subject's disease may influence both the intensity and outcome of search for exposure to the putative cause	When interviewer is not blinded
Attention bias	Subjects may systematically alter their behaviour when they know they are being observed	During food recording diets may (intentionally) be simplified, 'unhealthy' foods avoided or 'healthy' choices increased [225]

Berg et al [226] investigated selection and response bias is a dietary survey of Swedish children in fifth, seventh and ninth grades. They found significant differences between participants and non-participants with respect to socio-demographic and food variables, despite great efforts to obtain a high response rate. A decline in recorded foods during the recording period was also observed. They conclude that these two types of bias are likely to be present in dietary surveys involving children, and consequently this should be taken into account during the planning, analysis and interpretation of data.

Buzzard and Sievert [32] list identifying and minimising bias, particularly non-response bias and other sources of error as research priorities in dietary assessment methodology.

Randomisation, restriction, matching, stratification and statistical modeling are methods to control confounding.[28, 45]

Random error is the divergence, due to chance alone, of an observation on a sample from the true population value, leading to lack of precision in the measurement of an association. It cannot be entirely excluded, yet quality control procedures during each stage of the dietary assessment can increase the reliability and hence the precision.[45, 220] In general, individual biological variation, sampling error, and measurement error are the major sources of random error.[45] Apart from the above-mentioned quality control measures, adequate sample size or taking the average of multiple reference measurements (dietary recalls or records) per subject are the best ways to reduce random error in dietary surveys. This has recently been reviewed by Volatier et al.[166] Formulae for calculating sample size and repeat measurements are available, but cost considerations always play a role.[166]

Systematic and random errors may each occur at the intra- and inter-individual level, [222] the characteristics of which may be summarised as follows (based primarily on text provided by reference [222]):

TABLE 2.5: MEASUREMENT ERROR IN DIETARY DATA

	Random	Systematic
Intra-individual	• Reflects day-to-day variation above and below the individual's true long-term intake • Is the major source of error in dietary data • The magnitude varies by nutrient: Macronutrients vary less, because they make a large contribution to total energy intake; Micronutrients vary more because they are often concentrated in certain foods, and their intake is strongly influenced by food choices for the day • The effect of this error is to attenuate the strength of association, causing the correlation or regression coefficients to be biased toward zero • Methods are available to adjust for this error, provided replicate measures of diet are available. Examples include reliability ratio, correction factors, within-person variance, all of which can be described as “approximation of results that would otherwise be obtained if the estimates of long-term diet were available”	• Best typified by under and over-reporting of intake by some individuals (for example underreporting of energy and fat intake by overweight subjects) • Depends on the accuracy of the reported intakes by the subject or the interviewer and the detection of misreporting and especially underreporting [166]
Inter-individual	• Caused by using only a few measurements per subject in the presence of random within-person error	• Results from systematic within-subjects error that affects subjects non-randomly, for example using incorrect nutrient composition values for some foods may appear to affect all individuals in the same direction, but their impacts are not the same, since consumption of these foods is likely to differ among subjects • Biases in national representative dietary surveys can be linked to non-responders, since the non-responders may differ significantly from those participating in a study. Affected by control of missing or undefined data, the description of foods, procedures to code and aggregate single food items, and data check procedures [166]

Some researchers (for example reference [111]) avoid the use of the terms ‘systematic’ and ‘random’, since they reason that systematic errors can be randomly distributed. They propose a distinction between unbiased and biased methods, and differential and non-differential errors.

Based on this reasoning a measurement X' is defined as an unbiased measurement of X if the average measurement approaches the true measure as the sample size increases. Unbiased measurements result for $X' = X + \varepsilon$, where ε is a random error variable with expectation 0. On the other hand, X' is a *biased measurement* when the average measurement does not approximate the true intake, and ε from the equation does not equal 0.[111]

Both biased and unbiased errors can be either differential or non-differential. Measurement error is *non-differential* as long as the error distribution is identical for all individuals of a study or for each subgroup of a population. In the above equation this would occur if the distribution of the error variable ε would be the same for every individual under study. Measurement errors are *differential* if the participants of a study react differently to a measurement method that is used within a study.[111] For example, if hospital patients report a lower variance in their diets than controls, this may be reflective of their current diets, but not the true variance of their usual diet - unbiased, but differential error - because the variance and therefore the errors are differential. If, however, obese individuals underreport their fat intakes, whereas lean subjects report accurately, then this would constitute biased differential error.[111, 122]

The distinction between differential and non-differential measurement error is important for error assessment, adjustment, and correction strategies.[111] For non-differential errors the direction of influence on an estimated exposure-disease relationship is presumed to be biased toward zero. Differential errors influence exposure-disease relations in ways that can only be predicted if information about error in all subgroups under study is available.[111, 122]

It is not possible to statistically distinguish random errors from true intra-subject variation as described in the introduction,[220] unless replicate measures of diet are collected and these replicates are time independent, meaning that the replicates are taken on at random, preferably non-consecutive days.[222] Alternatively, observed intra-subject variation represents the sum of true variation plus the remaining sources of random measurement errors.[220]

The goal of a study or the aim of a dietary assessment tool will dictate the required accuracy. Equally, the impact of measurement errors on the design of a dietary assessment method depends on the aims to be achieved by the study. For example, the larger the random error, the greater the size of the sample required for estimating mean nutrient intake of a group. Also, increased random error increases the number of replicate measurement days necessary to define the distribution of usual nutrient intake of an individual.[111, 220]

The sources of error related to data collection and recording are of prime interest in this study. Intake is recorded as foods. On this level of error (also called ‘misrepresentation’, ‘false memory’, ‘misreporting’ or ‘distortion’), [91, 94, 95] the following can be distinguished:

- *Phantom foods* are items reported but not observed eaten. These are also called 'intrusions', 'commissions' or 'false positives'.
- *Omissions* are items not reported eaten but observed eaten, also called 'false negatives'.
- *Elaboration* is the process of unintentionally distorting information, for example when the participant uses general ('generic') knowledge of his / her diet to substitute for the true past information. The phenomenon is also called 'substitution'.
- *Matches* are foods reported eaten and observed.

If the participant accurately remembers eating a food item, or erroneously assumes that it was eaten, then the quantity consumed must be estimated. This information may not be salient for most respondents. Estimating usual amount is a complex cognitive task, since:

- Food frequency and portion sizes are not constant over time.
- Inferences and calculations must be made when the respondent's frequency and portion size differ from those given.
- Respondents often do not pay attention to frequencies and portion size when eating.
- Respondents do not have clear mental images of portion sizes eaten when estimating consumption.[88]

Error can also be associated with the compilation of nutrient data. The following is of importance in this regard: [27]

- Inadequate sampling procedures
- Inappropriate analytical methods
- Errors in analytical methods
- Lack of standardised conversion factors for calculating energy and protein content
- Inconsistent terminology
- Incorrect description of individual food items
- Inconsistencies from genetic, environmental, food preparation and processing factors

Kris-Etherton & Etherton [227] and Stumbo [228] emphasised the importance of a current nutrient database, since new foods enter the marketplace and existing foods undergo compositional changes, particularly in the case of fatty acids. In South Africa it was found that fatty acid intake from seven-day dietary records differed when the 1986 Medical Research Council Food Composition Tables were used for analysis compared to the 2000 SAFOODS.[229]

From the above it is apparent, that variability and error are inherent to dietary assessment. Any attempt to evaluate dietary intake needs to take this into consideration. Since underreporting of energy intake appears to be the most common form of misreporting, an analytic framework of predictors of accuracy of reporting has been proposed by Tooze et al (Figure 2.3).[230]

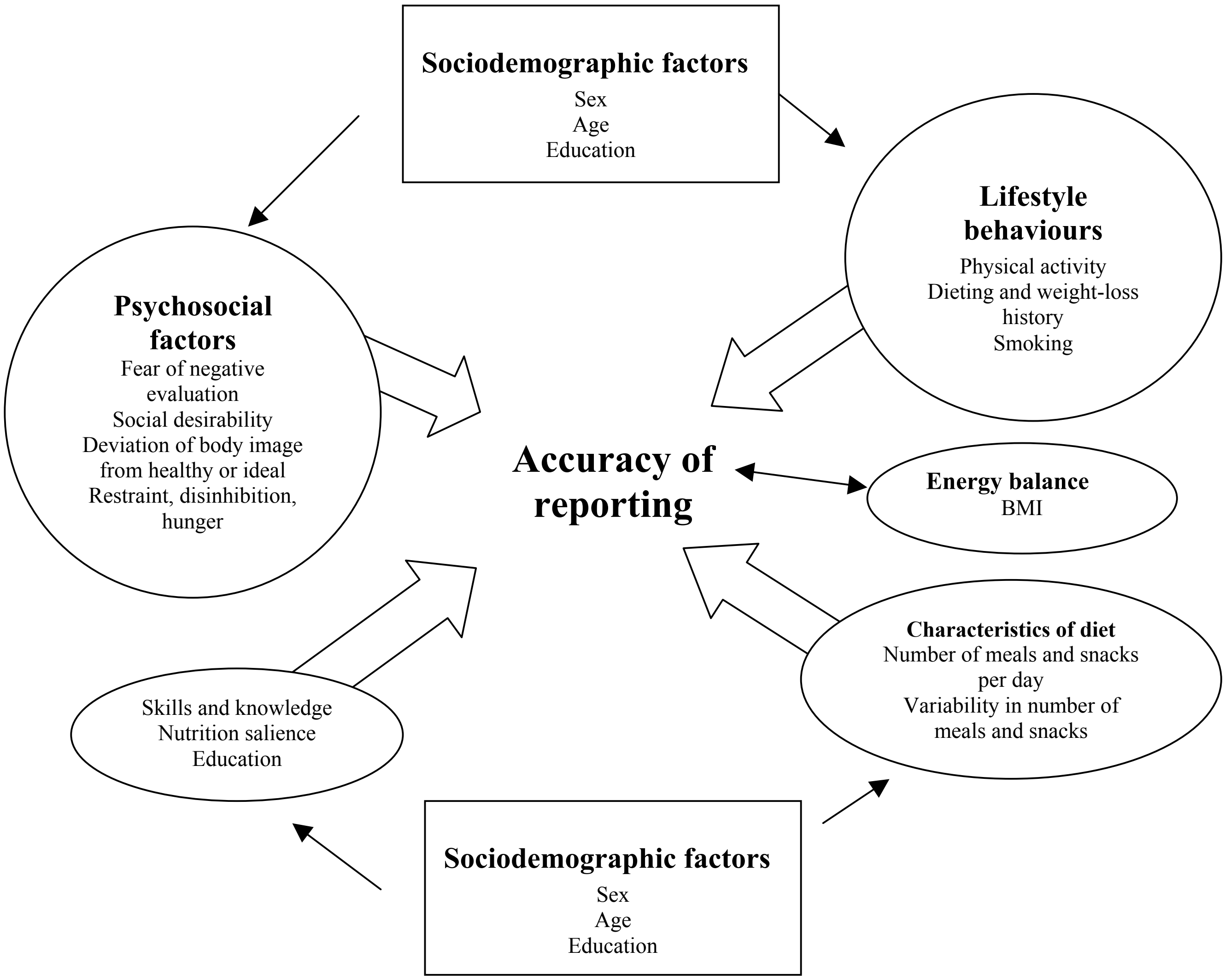

FIGURE 2.3: ANALYTIC FRAMEWORK OF UNDERREPORTING OF ENERGY INTAKE (from reference [213])

2.4.2 Reliabilty

2.4.2.1 Definition

Reliability of measurement (or the lack of it) is the extent of unsystematic variation in the quantitative description of some characteristic of an individual when the individual is measured a number of times.[231] It is a central consideration of validity concerning the process of data collection.

Marais and Mouton [232] as well as Miller and Achterberg [233] state that reliability refers to the requirement that the application of a valid measuring instrument to different groups under different sets of circumstances, should lead to the same observations. Thus reliability could be defined by posing the following question: "Will the same methods used by different researchers and/or at different times produce the same results?" even though Joachim [221] has argued that the term does not have a universal meaning to all researchers, and that the definition "stability of an instrument and measurement process when it is applied under standard conditions" may be problematic in dietary studies, as the meaning and relevance of 'standard conditions' are not clear. They prefer to define reliability as "the ability of an instrument to consistently measure what it aims to measure".

Wilkin et al [182] state that the more reliable a measure is, the lower the element of random error. Unreliable measures cause problems, when the aim of the tool includes the following: [231]

- Comparison or ranking among individuals by means of the measure
- Assigning individuals to groups based on scores obtained in the measure
- Prediction of other traits
- Assessing the (systematic) effects of other factors on the measure

2.4.2.2 Types of reliability

A number of types of reliability are, according to Wilkin et al,[182] of importance in the assessment of instruments:

(i) Consistency over time (Test-retest reliability)

This type of reliability is also called 'reproducibility' [80] or test-retest 'stability',[231, 233] and refers to the ability of an instrument to produce the same estimate on two occasions, assuming nothing has changed in the interim.

Thus, the measure is administered twice to the same group in a relatively short time, under the same conditions.[182, 184] This will determine

- whether respondents interpret questions in the same way during the first and second administration and
- whether the tool is 'stable' (that is answers remain the same despite mood changes or other recent, unrelated events).[184]

Wilkin et al [182] have identified two problems with this approach:

- There may have been real changes in the population between the two administrations. The magnitude of this problem depends on the length of time which lapses between administrations, and the stability of the variable being measured. Consequently, although the results from two separate dietary assessments disagree, the method may not be imprecise: the food intakes may indeed have changed. [27]
- Subjects may either undergo a learning process or remember the responses they gave at the first administration. Statistical techniques are available to overcome this problem, but this may lead to rejection of a very Afine-tuned@ instrument which reflects responsiveness in measure rather than random errors. Wilkin et al [182] and Ghiselli et al [231] label this a 'carry-over' effect, which may change the true score or create preservation effects.

Reproducibility provides a partial answer to the question of validity. Furthermore, reproducibility studies can uncover problems in instrument design, respondent instructions, or quality control. Finally, once the reproducibility of a tool is known, that information can be used to judge between the administrations (for example for monitoring or impact assessment).[80] In addition, knowledge about the reproducibility of an instrument can be used to increase the power of a study.

Some researchers [140, 221] distinguish between data and the database. The structure of the database is the format, or list, or questions used to collect data. The data are the information collected from the subjects, and the database is the total information collected using the structure of the questionnaire. Similarly, dietary studies consist of three components: the input, the data collection questionnaire / tool / instrument, and the compiled data / database.

Using the above as starting point, some researchers reason that a structure (that is a data collection tool) cannot per se be tested for reliability, reproducibility or validity. Consequently, reproducibility of data is separated from reliability and validity of data.[221]

Joachim [221], for example, argues that reproducibility (like reliability and validity) is a logical operator that can be assigned the value of true and false. Since data can be reproduced twice in dietary studies - when the data are deemed to be reliable (that is reflecting what it should) and when the data are deemed unreliable (giving the same, but wrong result) - she claims that there is no correlation between the reproducibility and validity of data. She proposes the following mathematical approach, in which reproducibility is seen as a function of specific variables or groups of variables. In this context, reproducibility is affected by three input variables and is described as follows:

Data collected are f(x1, x2, x3 xn)

Therefore, f(person, place, time)

A study could be deemed reproducible of all these variables are pegged. Any change in the variables will result in a change in the degree of reproducibility. The food intake of a *person* (that is the subject) depends on socio-economic status, age, education, and ethnicity. The geographic location influences food availability and price (that is the *place* variable in the relationship). Seasons, trends and the length of the study, form part of the *time* factor.

Gibson [27] uses the term 'precision' when a dietary assessment technique gives very similar results when used repeatedly in the same situation. The precision is seen as a function of the measurement errors and the true variation in intakes (see introductory sections). In addition, the precision of a particular dietary assessment technique depends on the time frame of the technique, the population group under study, the nutrient of interest, the technique used to quantify foods, and the inter- and intra-subject variability.[27]

For quantification purposes, test-retest reliability can be expressed in terms of a correlation coefficient, the so-called coefficient of stability. Coefficients >0.7 are considered to be a minimum standard of stability.[233]

(ii) Consistency between different users (Inter-rater reliability)

This type of reliability investigates the judgment by a rater, and is being tested when two different raters are required to form an opinion of the same measure. The Kappa coefficient of agreement is used to statistically assess the probability of chance agreement. This coefficient can range from -1 to +1, where -1 indicates agreement worse than chance.[182] Alternatively Leedy [234] states that an agreement of 85% is acceptable and indicates the number of items out of the whole set about which the observers agreed. The major source of error with this method is inconsistency between data collectors, usually due to inadequate training. Training is necessary to ensure that all factors relating to the administration of the tools are kept constant.

In the context of nutritional assessment McCall and Cotton [235] determined the inter-rater reliability by comparing results obtained by a dietitian to those of a nurse. They calculated an agreement confidence interval (above 95%) and Kendall's Coefficient of Concordance as indicators of inter-rater reliability. Low inter-observer variation has been highlighted by Kondrup et al [36] and Jones[185] as very important features of reliability of nutrition screening tools.

(iii) Internal consistency

Internal consistency refers to the extent to which all items in a scale measure the same dimension. Statistically it can be seen as an estimate of homogeneity, during which the extent to which individual items are correlated with each other and with overall scale scores are determined. The Cronbach alpha and Kuder Richardson statistics are examples of appropriate statistical approaches in this context. It must be kept in mind that this analysis would only be relevant to measures containing items relating to one dimension.[182] Using this approach, the instrument is administered to the sample once. The Kuder-Richardson formula is used when items are scored dichotomously, whereas the Cronbach coefficient alpha is used when several answers are possible.[233]

Keller et al [184] recommend that this analysis be performed during pretesting of a tool in a developmental sample (which can be the same as that for the initial validation study). This requires a diverse population of subjects with varying intakes. A Cronbach alpha of 0.7 indicates that all of the items contribute to the discriminating power of the scale and there is good internal reliability. Lower values indicate an excess of nuisance items, or too few items in the scale. Values >7 indicate that there are items measuring the same thing, and that some are thus unnecessary.

McCall and Cotton [235] investigated the internal consistency of a nutritional assessment tool: They established the dimensional structure by submitting data to a correlation matrix and principal component factor analysis with varimax rotation. They reasoned that this approach would answer the question “Does the tool ask the right questions or is there some overlap?” High agreement (overlap) was interpreted as that there was overlap (duplication) of areas covered by the questions, while low internal consistency was associated with a multi-faceted nature of items. A similar approach was followed by Johnson et al.[236]

(iv) Equivalent forms reliability

This fourth type of reliability refers to the extent to which two different versions of the same instrument (for example Form A and Form B of a scholastic aptitude test) yield similar results.[234]

2.4.2.3 Measurement of reliability

There is no universal test of reliability. In addition to the above specific types of reliability and the ways to measure them, Mouton and Marais [232] as well as Ghiselli et al [231] mention the use of parallel forms of a test and the split-half method (comparative parts of a test) as possibilities within (psychometric) test construction. In the case of the latter, this would mean that the items in the scale (which are supposed to measure the same attribute) are (randomly) assigned to two different sets. Each set of items should correspond in the way they classify subjects in the study.[231, 237] Furthermore, it may be necessary to re-establish reliability in differing conditions.[182]

In statistical terms, the most common indices to quantitatively describe degree of reliability of measurement are the standard error of measurement and the reliability coefficient. [231] In respect of the latter, Wilkin et al [182] claim that, in general, the accepted reliability standard is 0.5, since random error will then tend to average out in large samples.

Where statistical techniques are inadequate or inappropriate to numerically check for reliability, standardisation and strict quality control measures during data-collection remain the key requirements for reliability. This means focusing on the four areas in which threats to reliability can emerge (see section "Factors influencing validity and reliability").

2.4.3 Validity

2.4.3.1 Definition and principles

Historically, the most common definition of validity is: "A measurement procedure is valid if it measures what it purports to measure." [238] Within the psychometric and education context, validity, more specifically, refers to the appropriateness of inferences from test scores or other form of assessment.[231] Simply stated, this would mean the following: Given a research question or an aim, how useful (that is valid) are the answers (that is the information) provided by the test score? Thus a valid measuring instrument can be described as measuring what it is supposed to measure and as yielding scores whose differences reflect the true differences of the variable being measured rather than random or constant errors.[183] Beaglehole et al [45] state that a study is valid if its results correspond to the truth with no systematic errors and the random error as small as possible.

All of the above definitions require that the truth be known.[50]

Frongillo [239] has stated that validation is the process of determining whether a method is suitable for providing useful analytical measurement for a given purpose and context. He claims that all of the following criteria must be fulfilled for a method to be called valid for a particular purpose and context:

- Its construction is well-grounded in an understanding of the phenomenon
- Its performance is consistent with that understanding
- It is precise within specified performance standards
- It is dependable within specified performance standards
- It is accurate within specified performance standards
- Its accuracy is attributable to the well-grounded understanding for that purpose and context

More than 40 years ago, Becker et al [238] stated that when the researcher has a (i) perfectly calibrated tool and a (ii) purely objective technique for its use, the error variations tend to be at a minimum. However, nutritionists are faced with a far more difficult problem in accurate data collection:

There is no ultimate criterion-measuring device that can be used for the calibration of other devices (no 'golden standard'), and very often no established criteria exist. Consequently comparative validity, which poses special challenges, judgement and logical and empirical processes [240] is the only alternative, since validating 'usual' or 'habitual' intake presents overwhelming practical difficulties or is actually impossible.[50]

Buzzard and Sievert [32] state that a calibration study involves the collection of dietary data from a subset of study participants by using two different dietary assessment methods. It is the comparison of one method of dietary assessment to another with the aim of better understanding the level of agreement or relationship between the two methods. There is much overlap in the nature of the studies to which the principle is applied. Typically a less-detailed method (test method) is compared to a more detailed method, which is assumed to provide more accurate estimates of intake. The objective of the calibration study is to quantify the bias of the less-detailed method in relation to the more detailed method. This permits adjustment (calibration) at the group level of intakes derived from the less detailed method. Livingstone and Black [159] state that calibration studies are studies of relative validity to distinguish them from studies of validity that use external markers of intake. Given that a gold standard for validation is not available, the best one can do is to calibrate one method against another believed to be more accurate for the purpose at hand.[241] Since validation is hardly possible, calibration is a useful alternative, including comparison of results between studies, between methods and over time. Calibration is the measurement of the distance between measurements from two different instruments, or the measure of change in the accuracy of measure of an instrument over time.[241]

From the above it is evident that the terminology around validation and calibration in the nutrition context has not been standardized.

2.4.3.2 Types of validity

Since the validity of an instrument is a function of the specific aim it is intended to achieve, it follows that different types of validation or validity evidence can be obtained. Thus, questions about validity cannot be separated from a consideration of the specific purpose for which a test is to be used. Equally, a test may have several purposes, which can vary in kind and scope. Consequently a given test may have a moderate validity in achieving one aim, yet have good or poor validity in another respect.[231] Typical examples from the nutritional epidemiology literature emerge where a test might exhibit a different validity in respect of quantitative precision, versus classification agreement, versus ranking of individuals or establishing prevalence in groups.

In clinimetry, which focuses on the quality of clinical measurement, where quality includes both the quality of the measurement instrument and the quality of performance of the actual measurements, [242] this has lead to the convention of referring to different types or forms of validity, each of which is important in different situations.[183, 234, 240]

Face, content, representative and/or consensual validity, as a group, refer to the overall relevance, adequacy and / or (relative) representativeness of the components of an instrument, as judged by content experts and / or potential users of the tool.[36, 233, 234, 240, 242] Sometimes a differentiation between the terms is made, but, generally speaking, these types of validity are related, are based on judgment and tend to be subjective.[240, 242]

Criterion-related validities range from situations where an external ‘gold standard’ (that is the criterion which reflects ‘truth’) is available, to relative validity where either the truth is unknown or not (yet) measurable.[183, 193, 231, 240, 242] Establishing these forms of validity tends to be more rigorous than in the above-mentioned group. It has been claimed that these validities (particularly criterion validity) put validation “on the road to good science.” [243] An example of this most powerful approach in the nutrition context is the use of doubly labeled water as the criterion for energy expenditure. In the case of relative validity, the measurement obtained by the test method is compared to the results from another method or outcome variable assumed to be more accurate or indicative of the truth. Examples are *construct validity*, where the truth is a trait hypothesised to exist, but there is not one real-world counterpart for it.[122, 183, 231] In this context *convergent validity* refers to the agreement or correlation of independent measurements that are theoretically or logically related.[233, 238, 240, 244] *Discriminant validity*, on the other hand, is inferred when a measurement of a construct successfully discriminates between people known to have differing amounts of the trait being represented by the construct.[233, 244, 245] *Predictive validity* is a type of criterion-related validity that refers to the accuracy with which future outcomes (for example growth in the form of weight for height, weight gain) is forecasted by the test method.[231, 238, 240, 246] A substitute for predictive validity can be *concurrent validity*. In this case, another, currently present trait is measured in the place of the future outcome. From the above, it is clear, that construct validity is not an aspect of validity that is exclusive to other types of validity. In the nutrition literature some of these types of criterion-related validities have been used interchangeably (for example references [198, 208, 235]) or have been applied differently (for example when validating fruit and vegetable intake the use of biomarkers was called criterion validity by Murphy et al [208] and construct validity by Bodner et al (1998).

2.4.3.3 Validity of (dietary) screening tools

A screening tool is valid if it correctly categorises people into groups with and without disease, as measured by its sensitivity and specificity.[45] Caan et al [49] have consequently suggested that sensitivity and specificity are the best indicators of validity of a method for dichotomous classification because of the ability of these parameters to generalise results to populations amongst whom the prevalence of the phenomenon varies markedly. Both, sensitivity and specificity are thus descriptors of the accuracy of a test.

Sensitivity is the proportion of truly ill people in the screened population who are identified by the screening test.[45] The greater the sensitivity of a test, the more likely that the test will detect persons with the condition of interest. Thus, sensitivity is measured in the group of subjects who test positive by the reference method or 'golden standard' and reflects the true positive rate.

Specificity is the proportion of truly healthy people who are identified by the screening test. [45] The greater the specificity, the more likely that persons without the condition of interest will be excluded by the test. In a group of subjects who test negative on the reference method, specificity is defined as the true negative rate in the screening procedure.

A high sensitivity for a screener may give false positives, with more subjects classified as having a high fat intake than is actually the case. Equally, a high screening specificity may give false negatives (based on reference [247]).

Thus having both, high sensitivity and high specificity, represents the ideal. However, a balance must usually be struck between the two, because the cut-point between normal and abnormal is usually arbitrary, and because very often sensitivity and specificity are inversely related. Apart from the inherent aim of the screener, availability of funds and resources to support interventions, the seriousness of the disease, the distribution of the risk factor as well as local experience of the severity of risk, are amongst the factors that will determine whether high specificity or high sensitivity will be favoured in a particular situation. In general, raising a threshold for considering a result to be positive typically will lead to a gain in specificity (fewer false positives) but a loss in sensitivity (more false negatives or missed cases). On the other hand, lowering the threshold for considering a result to be positive typically will reduce the level of false-negatives (raise sensitivity) and increase the likelihood of false-positives (lower specificity). Very specific tests are often used to confirm the presence of a condition.[45]

In conjunction to the above, *predictive value* affects the usefulness of a screener. Predictive values depend on sensitivity and specificity, but most importantly, on the prevalence of the condition in the population tested or the pre-test probability that a subject has the condition of interest. Positive and negative predictive values can be distinguished, where positive predictive value is defined as the percentage of persons with positive test results who actually have the condition of interest, and negative predictive value is the probability of the condition being absent if the test is negative. Thus these two measures address the estimation of probability of disease or a specific condition of interest (for example high fat intake).[45]

Relative risk (risk ratio) is the ratio of the risk of occurrence of a disease among exposed people to that among the unexposed whilst the *odds ratio* is the ratio of the odds of exposure among cases to the odds in favour of exposure among controls.[45]

2.4.4 Validation studies

2.4.4.1 Background

Comparative validation is not new: As early as 1942 Huenemann and Turner published an exemplary validation study in children aged six to 14 years, where, at the beginning, they obtained a detailed diet history from each of the subjects. This was followed by a ten to 14 day precise weighed diet record, which was repeated every three to four months, three or four times, so that the period of time covered for each child ranged from at least six months to one year. Based on the variation of the amount of nutrients, the authors concluded that no single diet record could be considered 'typical' of an individual subject's food intake.[238]

An early review of dietary intake methodologies and validation studies was compiled by Becker et al [238]. In the 1980's various additional reviews on dietary validations were published.[50,77, 248, 249] In the 1990's Friedenreich [91] reviewed methods that measure past diet and Gibson [171] wrote a general review about dietary assessment. Jones [185] specifically critiqued dietary assessment methodology. For the past few years a register for dietary assessment calibration and validation studies has been available online (www-dacv.ims.nci.nih.gov/).

2.4.4.2 Validation studies in children

As mentioned before, Mc Pherson et al [25] published a review of validation studies in school-aged children. The following discussion is thus limited to studies focussing on validation of screeners and studies not discussed previously (under FFQ or food records) or mentioned in the McPherson et al review.

A seven-item fruit and vegetable FFQ had a low validity among third-grade students when compared to seven-day food records.[250] The major problem was the severe overestimation by the FFQ. Cognitive problems were offered as main reason. Field et al [195] compared four brief questionnaires for measuring fruit and vegetable intake with estimates from three 24-hour recalls on non-consecutive days in adolescent. They found the short methods useful for ranking but not for estimating prevalence of consumption of five or more servings of fruit or vegetables per day. The validity and reproducibility of a questionnaire aimed at assessing fruit and vegetable intake was evaluated in sixth grade Norwegian children. It was compared to seven-day food diaries. Reproducibility was acceptable, as was the comparative validity of vegetable intake. Fruit intake, however, was overestimated.[251]

Baranowski et al [252] assessed the validity of a 'Food Intake Recording Software System' against observation of school lunch and a 24-hour recall. They concluded that this lower-cost approach was promising, though somewhat less accurate than the 24-hour recall.

Jonsson and Gummeson [226] assessed reliability and construct validity of a method that utilised picture stacking to measure food choices (milk, margarine, bread, cereals) for breakfast. They reported that for milk and margarine reliability and construct validity were good, but random error or a trend towards healthier choices played a role in the other cases. Pictures were also used for adults in the Cardiovascular Health Study. Food frequency scores were obtained from a picture sort procedure, which yielded relative validity similar to conventional FFQ.[253]

'Yesterday's Food Choices' is a 33-item instrument validated for American Indian children in fifth to seventh grade.[92] A modified diet record-assisted 24-hour recall was validated by direct observation among third-grade American Indian children. Weber et al [254] concluded that at group level the reported macronutrient proportions of total energy intake were accurate.

As part of the Child and Adolescent Trial for Cardiovascular Health (CATCH) a short Food Checklist (CFC) was developed as a measure of PFE, PSFE and sodium intake in middle school students. Children (n=365 seventh graders) provided yes / no responses with respect to intake on the previous day for 40 items on the checklist. Results were compared to 24-hour recalls and reproducibility and validity were demonstrated.[255]

Habitual meal patterns and intake of foods, energy and nutrients in 15-16 year old Swedish girls was measured with a 'diet history'. Seven-day food records served as reference method. The former was found to perform as well as the reference method in terms of classification agreement of meal patterns. Also energy and nutrient intakes were similar. For individual foods there was less similarity.[256]

In Pretoria, South Africa, a modified diet history was compared to seven-day precise weighing food records in six to eleven-year old white children. It was concluded that the shorter and more practical modified diet history gave results at least as satisfactory as the laborious and time-consuming seven-day precise weighing.[63]

Potgieter and Fellingham compared a 24-hour weighing method with a seven-day weighing method in black, Indian and coloured children in Pretoria. They concluded that there was no serious bias in the 24-hour weighing method and that it could be a rough estimate of population means.[173]

2.4.5 Factors influencing validity and reliability

Mouton and Marais [232] identified four major variables to keep in mind when attempting to ensure that validity and reliability are not threatened. Some of these fall in the cognitive perspective, whilst others could be classified as relating to the situational perspective.[257]

- *The researcher / interviewer/field worker* is the first factor mentioned by Mouton and Marais.[232] In this regard the researcher's characteristics such as affiliation, image and distance from the participants, as well as his/her orientations such as bias-producing cognitive factors, attitude structure expectations and role expectations can play a role. Referring to measurement of diet, Gibson [220] specifically mentions the use of incorrect questions, incorrect recording of responses, intentional omissions, biases associated with the interview setting, distractions, confidentiality and anonymity of the respondent, and the degree of rapport between interviewer and the respondent. In the Bogalusa study Frank et al [258] showed that interviewer recording practices had an effect on the recorded nutrient intakes of children.

- The individual who participates in the research project (*participant / respondent / subject*) who, in the so-called guinea-pig effect can show signs of memory decay, omniscience or interview saturation is the second factor highlighted by Mouton and Marais.[232] They add the perceived role, level of motivation and response patterns as being participant orientations that can also influence validity and reliability. Leedy [234] refers to this phenomenon as the reactivity or Hawthorne effect, and states that it specifically is a threat to the internal validity of a study. Diet-related examples given by Gibson [220] in this regard include over-reporting of ‘good’ foods such as fruits and vegetables and under-reporting of ‘bad’ food such as fast foods and alcohol, leading to a so-called prestige bias. Memory lapses, like forgetting to report the ‘minor’ parts of a meal (for example dressings), inability to report portion sizes, and the so-called flat slope syndrome, whereby respondents tend to overestimate low intakes and underestimate high intakes, are further examples.[220] Whilst respondents’ inability to estimate their intake reliably is an important factor influencing reproducibility, Block and Hartman [80] state that methodological explanations are more likely to play a role.

- Mouton and Marais [232] have listed the *measuring tool* (questionnaire / interview schedule) as a third factor which affects validity and reliability. They identified question sequence, open / closed questions, ‘don't know’, mid-position selection, questionnaire length, item sensitivity, leading questions and fictitious attitudes as aspects of importance. The sources of measurement error identified by Gibson, [220] that could be classified under this heading, include coding and computation errors (as when 'standard / reference'quantities of intake are do not reflect the intake of the subject and when these intakes are incorrectly converted to grams eaten). Another source of error can be found in the compilation of nutrient composition data (which can be random, systematic or true [like geographic / seasonal] variability, or due to errors in the nutrient analysis of food items, or the compilation of the computerised data base). An important factor affecting specifically the reproducibility of a tool is the variability it permits.[80] An instrument which does not include portion sizes, or which has limited response categories about frequency of consumption, is likely to have a higher reproducibility score, because it allows less variability. In such a case high level of reproducibility is desirable, but not sufficient to ensure validity. The physical questionnaire design (for example layout) and instructions given to subjects can also affect reproducibility.[80]

- The final factor listed by Mouton and Marais [232] is the *research context* (broad or specific spatio-temporal circumstances). This refers to time, cultural and political factors as well as the research setting as such. Leedy [234] illustrates this threat to internal validity in terms of subject selection, for example the use of volunteers and convenience sampling, and calls it 'experimenter expectancy' which may lead to a selection bias. Within the dietary assessment context, Block and Hartman [80] point out that reproducibility is clearly influenced by the elapsed time between two administrations.

2.4.6 Implications

Validity in nutritional assessment is not a 'black and white' issue: Firstly, because no criterion exists, the focus is on relative or comparative validity and varying degrees of validity are observed. Secondly, validity is dependent on the population and the context. Thirdly, it is important to differentiate between the validity of the measurement instrument and the actual performance of the measurement. If the measurement is performed sub-optimally, the instrument may be sufficiently valid, but the performance may not.[242]

In general, Leedy [234] lists four possible precautions that can help to enhance the internal validity of a study: Controlled laboratory settings, double blind experiments, unobtrusive measures and triangulation. For qualitative research, strategies that can be added include: spending extensive time in the field, performing negative case analysis, obtaining feedback from others and respondent validation. On the other hand, in order to improve the external validity, real-life settings, the use of representative samples and replication in a different context can be considered. Practical and ethical considerations usually necessitate compromise.

As far as reliability is concerned, Leedy [234] stresses consistent administration of instruments. This implies standardisation from one situation or person to the next. Secondly where judgments are required, specific criteria should be established to indicate the kinds of judgments that must be made. Finally any research assistants who are using the tool should be well trained so that they obtain similar results.

In respect of dietary assessment Gibson [220] and Kohlmeier [259] point to the following practical implications:

Quality control needs to be implemented at each stage of the dietary assessment. Quality control refers to the range of procedures undertaken during data collection and analysis to ensure quality of measurement. This involves steps to prevent, reduce, detect and correct errors. The following aspects deserve special attention:

- Researcher and field workers:
  - Training and retraining for interviewers and coders, referring to aspects such as the extent of probing and use of probing aids, wording of questions, participation of other persons in data collection [260]
  - Standardisation of interviewing techniques and questionnaires
  - Pre-testing of questionnaires
  - Pilot surveys
  - Training of interviewers to anticipate and recognise potential sources of distortion and bias
  - Minimise non-response by training interviewers to convey understanding, trust and warmth
  - Concentrate at avoiding value judgments
- Respondent:
  - Implement knowledge about cognitive processes involved in diet recall, specifically to improve question comprehension, improving information retrieval, improving estimation of quantities, improving response formulation [91]
  - Attention to memory by using probes, visual aids et cetera
- Data handling and computer program:
  - Credibility of software and nutrient database

  - Reduction of number of steps in data processing
  - Duplicate entry for a certain percentage of observations
  - Programming for error detection (frequency distributions, flagging et cetera) [260]
  - 'Coding rules' to deal with incomplete or ambiguous food descriptions / meal codes
- Sampling:
- Collection of supplementary information

In conclusion, from the review of the literature it is evident that the dietary habits of children are very often not in line with international recommendations, particularly in respect of fat intake and for reducing risk of CNCD. Dietary assessment of children can take on many forms, but no one method is perfect. Screening appears to be an attractive alternative, but, as in the case with detailed assessments, the comparative validity of these methods must be established in the population for which the tool is intended.