Validation of a multi-nutrient food frequency questionnaire to determine nutrient intakes of New Zealand toddlers 12-24 months old

Emily Olivia Watson

A thesis submitted for the partial fulfilment of the degree of

Master of Science

at the University of Otago, Dunedin, New Zealand

December 2012

Abstract

Adult health problems such as obesity and diabetes have been associated with dietary intake during the pre-school years. However, due to the difficulties involved in measuring the diet of toddlers, limited data is available on their dietary intake. A food frequency questionnaire is a cost effective and easily administered dietary assessment method and is the preferred choice in many epidemiological studies. As there is no such food frequency questionnaire available for use in New Zealand toddlers, the aim of this study was to determine the ability of a 97-item food frequency questionnaire to assess the usual dietary intake over the past four weeks of New Zealand toddlers aged 12 to 24 months.

Participants were the primary caregivers and their child. Primary caregivers completed two administrations of the food frequency questionnaire for their child, approximately four weeks apart. A five-day non-consecutive weighed diet record was completed over the four weeks between questionnaires. Relative validity was assessed in 153 participants by randomly selecting either their first or second food frequency questionnaire, and comparing nutrient intakes obtained from the food frequency questionnaire with those from the diet record. Spearman correlation coefficients, cross-classification and Bland-Altman statistics were calculated. Reproducibility was assessed in 152 participants, by comparing the two administrations of the food frequency questionnaire using intraclass correlations. Validity and reproducibility were assessed for both unadjusted data, and data that was adjusted for overall fruit and vegetable intake, using a single cross-check question, as used in the Prevention of Overweight in Infancy Study (POI-adjustment) or two cross-check questions as used in the Eating Assessment in Toddlers Study (EAT-adjustment), as recommended by expert groups.

Spearman's correlation coefficients for nutrients ranged from 0.37 for total fat to 0.66 for calcium, with a mean coefficient of 0.50. Coefficients increased minimally when the fruit and vegetable sections were adjusted for total intake. On average, 40% of participants were correctly classified into the same quartile, and only 4% were grossly misclassified. Bland-Altman statistics demonstrated that the food frequency questionnaire overestimated intakes, with mean unadjusted nutrient data being 28-78% higher than the corresponding diet record data, and mean POI adjusted intakes 12-60% higher. The single exception was

vitamin C which was not significantly different from the diet record. The food frequency questionnaire exhibited good reproducibility between administrations, with POI adjusted intraclass correlations ranging from 0.65 for vitamin C to 0.75 for calcium.

The food frequency questionnaire showed adequate to good relative validity for ranking participants according to nutrient intake using POI adjusted data, and high reproducibility. Our results compare favourably with previous literature, with validity and reproducibility similar, if not higher, than previous toddler multi-nutrient food frequency questionnaires. However, consistent with previous studies, the food frequency questionnaire was not suitable to assess absolute energy or nutrient intakes in individuals, overestimating the intake of virtually all nutrients.

Preface

This validation study was conducted at the Department of Human Nutrition, University of Otago, Dunedin, New Zealand. The food frequency questionnaire (FFQ) was developed previously by another student, based on a questionnaire used in 12-month old toddlers in the Southampton Women's Survey (SWS). Assessment of the adapted questionnaire's validity (as part of The Eating Assessment in Toddlers (EAT) Study) took place between February 2011 and June 2012 and was carried out by two MSc students, including this candidate. The candidate's supervisors, Dr Anne-Louise Heath, Associate Professor Rachael Taylor, and Dr Paula Skidmore were responsible for the concept and overall study design.

The candidate was responsible for the following under supervision:

- Applying for ethical approval
- Development of study protocols and study tracking sheets
- Procurement of study equipment
- Adaption of the study resources including the diet record and FFQ, and organisation of printing for the study
- Weighting and nutrient coding of the FFQ
- All advertising, recruitment, home-visits and data collection in Dunedin and Mid-Canterbury
- Advertising, recruitment, and data collection for an additional 36 participants in Wellington
- Entry of 45 5-day diet records, co-ordinating entry of remaining diet records and quality control processes for all other diet records.
- First data entry of all the FFQs, co-ordination of the second entry, quality control processes for the FFQs, and reconciliation of all data entry queries.
- All statistical analyses and interpretation presented in this thesis

Acknowledgments

I would like to express my gratitude to everyone who has helped and encouraged me during this thesis project.

In particular, I would like to thank:

My supervisors; Dr Anne-Louise Heath, Associate Professor Rachael Taylor, and Dr Paula Skidmore. Your encouragement, support and advice over the past two years has been invaluable. I never expected to have learnt so much and I am so grateful to have had the opportunity to work with you all.

Virginia, for recruiting the other half of participants, I would never have finished without you. For welcoming me to Wellington when I first moved there, your support, lengthy discussions and cups of tea were always what I needed to keep working through our project and get to the end.

Hayley Stevensen for helping with the diet record entry, and Ashleigh Barris for so efficiently entering and checking the diet records and questionnaires. Maha Hana, for so patiently showing me how to take the anthropometry measurements. Liz Fleming for answering my many Kai-culator related questions, and all your work getting us the final data. The University of Otago and the Human Nutrition Department for their generous financial assistance to complete the past year.

The wonderful families involved in this study, thank-you for putting the time and effort into taking part in this study, it would not have been possible without you. Also thank-you to all the participants who kindly convinced their friends, family, and colleagues to take part.

My family: Mum, Dad, Nana, Granddad, Sarah and Henry. Your support over the last two years, and every year before that, has been endless and I am extremely thankful. Thank-you for your unwavering support, for letting me explain everything to you, for pretending to understand, and offering advice you know I will ignore. To my wonderful group of friends, your moral support has made this thesis possible, I am lucky to have you all.

Table of Contents

1 **Introduction** 1

2 **Literature Review** 3

2.1 Introduction 3

2.2 Dietary Assessment Methods 3

2.2.1 Methods 3

2.2.2 Challenges of Dietary Assessment in Toddlers 6

2.3 Components of a Food Frequency Questionnaire 7

2.3.1 The Food list 7

2.3.2 Frequency Response Options 8

2.3.3 Portion Size Estimation 9

2.4 Designing a Validation Study 10

2.4.1 Study Population 11

2.4.2 Number of Participants and Replicate Days Required 11

2.4.3 Reference Method 12

2.4.4 Sequence of Administration 13

2.4.5 Nutrient Database 14

2.4.6 Statistical Analysis 15

2.5 Currently Validated Food Frequency Questionnaires for use in Toddlers 12 to 24 months old 18

3 **Methods** 26

3.1 Food Frequency Questionnaire Development 26

3.1.1 Modification 26

3.1.2 The Food List 26

3.1.3 The Frequency-response Options 27

3.1.4 Portion Size Estimation 27

3.1.5 Additional Questions 28

3.1.6 Expert Review and Pre-testing 29

3.1.7 Finalised EAT Food Frequency Questionnaire 29

3.1.8 Coding of the Food Frequency Questionnaire 29

3.2 The Validation Study 30

3.2.1 Ethical Approval 30

3.2.2 Recruitment and Study Population 30
3.2.3 Data Collection 33
3.2.4 Data Entry and Nutrient Calculations 37
3.2.5 Statistical Analysis 39
4 **Results** 41
4.1 Recruitment and Participants 41
4.2 Relative Validity 42
4.2.1 Energy and Nutrient Intakes 42
4.2.2 Correlation Coefficients 42
4.2.3 Bland-Altman Method 44
4.2.4 Cross-classification 45
4.3 Reproducibility 45
4.3.1 Energy and Nutrient Intakes 45
4.3.2 Correlation Coefficients 47
5 **Discussion** 48
5.1 Validity 48
5.2 Reproducibility 52
5.3 Study Strengths and Limitations 53
5.4 Future Research 59
6 **Conclusion** 60
7 **References** 61
8 **Appendices** 67

List of Tables

Table 2.1: Validity of food frequency questionnaires for assessing nutrient intake in toddlers 12-24 months old f23

Table 3.1: Recruitment activities in Dunedin, Mid-Canterbury, and Wellington f32

Table 4.1: Characteristics of the toddlers who completed the study and those who withdrew f42

Table 4.2: Average daily intakes of energy and selected nutrients in the 5DDR and FFQ f43

Table 4.3: Spearman correlation coefficients between the food frequency questionnaire and the 5-day diet record, compared with results from previous FFQ validation studies in toddlers.....43

Table 4.4: Bland-Altman statistics measuring strength of agreement for energy and nutrient intakes between the FFQ and 5DDR[a] f45

Table 4.5: Cross-classification of children by quartiles of energy and nutrient intakes based on the FFQ and 5DDR f46

Table 4.6: Average daily intakes of energy and selected nutrients for the first and second FFQ administrations 46

Table 4.7: Intraclass correlation coefficients for the repeated administrations of the FFQ, compared against the results reported by D'Ambrosio (2012) f48

List of Figures

Figure 3.1: Palm Volume; surface area of the palm and thickness.....28

Figure 3.2: Scanned image of a child's hand, with length and width measurements 35

List of Abbreviations

5DDR – 5-day diet record

CI – Confidence interval

EAT – Eating Assessment in Toddlers

FFQ – Food Frequency Questionnaire

MSc – Master of Science

NHANES – National Health and Nutrition Examination Survey

NNS – National Nutritional Survey

NZEO – New Zealand European and Other

POI – Prevention of Overweight in Infancy

SWS – Southampton Women's Survey

WHO – World Health Organisation

1 Introduction

Assessing the diet of pre-school children is becoming increasingly important as evidence accumulates that their diet does not only influence their current health but also has lasting health effects throughout the life course (1). There are no national data available in New Zealand on the health status of those aged under two, however the most recent New Zealand Health Survey found that 20.9% of children aged 2-14 years were overweight, and a further 8.3% were obese (2). This is a concerning situation, as childhood obesity is a predictor of adulthood obesity, which in turn is a risk factor for many non-communicable diseases including diabetes, hypertension, heart disease and some cancers (3). Nutrient inadequacy is also of concern due to its effect on a toddler's growth and development. Studies have found the prevalence of suboptimal iron status to be 33% and zinc deficiency 32% in New Zealand toddlers (4,5). There is a very limited amount of data available on the intake of other nutrients, making the risk of inadequacy relatively unknown.

The reason behind this lack of data in toddlers is in part explained by the difficulties encountered when trying to accurately measure their diet. They are too young to report their own food intake, and so information is required from a surrogate reporter. This can become more complicated if the child has multiple carers in control of their dietary intake (6). Another issue to consider is their high plate waste, and the difficulties carers have in quantifying the actual amount eaten. Toddlers often do not eat all the food offered to them and it requires an attentive caregiver to ensure all leftovers remain on the plate and are therefore able to be assessed. Dietary assessment methods also need to be mindful of the rapidly changing dietary patterns of toddlers. During infancy, a child's main source of nutrition is milk, however as they grow, more foods are introduced, and their diet diversifies. Their preferences can change quickly, and the study design must take this into consideration. It is therefore vital that the reproducibility of an FFQ designed for use in this age group is assessed. Another challenge, which is not limited to dietary assessment in the toddler age group, is the estimation of portion size. Participants find it notoriously difficult to accurately estimate portion sizes, with this estimation commonly attributed to be the greatest source of error in dietary assessment (7).

The food frequency questionnaire (FFQ) has low respondent burden, is relatively inexpensive, and assesses past usual dietary intake (8). Historically, FFQs have been validated only in adults, with more recent studies focusing on developing FFQs for use in older children and adolescents. More recently their use has been expanded to include the diet of toddlers, and therefore, questionnaires for use in this age group either need to be developed from scratch, or already available FFQs need to be adapted and revalidated. There are currently only six multi-nutrient FFQ studies which have assessed the validity of nutrient intake in toddlers between 12 and 24 months old (9–14), none of which have been validated in New Zealand.

Therefore, the objective of this study was to determine the relative validity and repeatability of the Eating Assessment in Toddlers (EAT) FFQ for assessing intake of selected nutrients in toddlers aged 12 to 24 months old.

2 Literature Review

2.1 Introduction

Relevant literature was found by searching the following databases: Medline via OvidSP from 1948 to October 2012, Google scholar and Science Direct up to October 2012, and by searching reference lists for relevant studies. Search terms used were: calibration, diet record, dietary assessment, food frequency questionnaire, infant, methods, portion size, questionnaire, reproducibility, surrogate, toddler, validation study, and validity.

2.2 Dietary Assessment Methods

The first recorded use of dietary assessment methods was in the 1920s, and since then they have been refined, modified and evaluated into the forms in use today (7).There are four main dietary assessment methods in use today: the diet record and 24-hour recall which assess actual food intake on specific days, and the dietary history and food frequency questionnaire (FFQ) which assess usual intake over a period of days.

2.2.1 Methods

2.2.1.1 Diet Record

A diet record involves the recording of all foods and drinks and the amount consumed, over a period of days. Detailed descriptions including the type of food, brand name, preparation methods, and portion sizes are recorded. Portion sizes can be estimated using household measurements and/or photographs provided by the study team. Ideally however, scales are provided, and the participants record the weight of each food and drink. Mixed dishes should be recorded by weighing and describing all raw ingredients, and recording the proportion of the total dish offered to the child.

Multiple weighed diet records are the most precise dietary assessment method for estimating usual dietary intake currently available (8). Nevertheless, for a diet record to be completed accurately the participants need to be highly motivated, numerate and literate. Weighing all foods eaten including those consumed outside of the home is time consuming and can be difficult, possibly leading to participants changing their eating habits to make recording easier (8).

2.2.1.2 Twenty-four hour recall

The twenty-four hour recall involves an interviewer asking the participant about their dietary intake either over the past 24-hours or the previous day. A standardised interview process has been developed which involves: a list of all the foods and beverages consumed, detailed descriptions of the brands and cooking methods and portion size eaten, and lastly a review of the recall to ensure all foods and drinks have been recorded (8). This is the method used in many national surveys, including the U.S. National Health and Nutrition Examination Survey (NHANES) and the New Zealand National Nutrition Survey (NNS) (8). Portion sizes are usually given in household units with answers aided by photographs, food models, or household tools such as spoons, bowls and cups. This structured approach is designed to retrieve as much information as possible from the participant, while minimising any errors in the data collection process (8).

A single recall by multiple participants can provide information on the intake of a population, however multiple recalls by each subject over a period of time is required to collect usual dietary intake (8). The interviewer needs to be highly trained to ensure enough detail is collected on each food including preparation methods, additions to meals such as seasonings, brand names, and to probe for any extra foods eaten and any leftovers. Computer programs are now available which provide the exact wording for the interviewers to follow, ensuring the interview process is standardised, and the required level of information is obtained (7).

2.2.1.3 Dietary History

The dietary history, originally developed by Burke in 1947, consists of three parts. The first is a detailed interview which gathers information on typical meals and snacks eaten throughout a typical day, including daily and seasonal variations of diet. The purpose here is to establish the usual pattern of dietary intake (15). The next two sections are a frequency questionnaire and a three-day food record using household measures. These two sections are used as a cross-check with the detailed interview to verify information on the kinds and amounts of foods eaten. The three part structure of the dietary history is rarely used today, with the third section often omitted completely (8). The dietary history provides information on the habitual dietary intake, however it is unsuitable for participants with erratic meal patterns, as it tends to underestimate any irregularities in food intake (16).

2.2.1.4 Food Frequency Questionnaire

The food frequency questionnaire (FFQ) is based on the checklist used in the dietary history and consists of two sections: a food list and a frequency-response section. A third section is often added asking about the usual portion size consumed, allowing quantitative data to be collected.

The FFQ has gradually been developed over the years to its current state as the principal method of dietary assessment in epidemiological studies, where their purpose is primarily to assess diet-disease relationships (7). Used for this purpose, FFQs are not necessarily required to report absolute intakes, or individual intakes, but instead to categorise or rank individuals, according to habitual dietary intake (17). The underlying principle of the FFQ is that average past diet, over a period of time, is the important exposure rather than intake on specific days. The FFQ is therefore designed to cover a time frame of usually one month to a year, depending on the study objectives (8).

The ability of the FFQ to assess habitual intake from one administration is a significant advantage over other methods. Dietary recalls and records require multiple days of data collection, whereas the FFQ requires only one administration usually lasting 15-30 minutes depending on the number of questions included (8). This single application reduces the burden on the participant, increasing completion rates in studies.

Other advantages of the FFQ include the relative speed, and cost-efficiency of data processing. The questionnaire does not need to be administered by a trained interviewer, and can be completed at home. This allows the questionnaire to be posted to participants, increasing the ability for data collection over a greater geographical area. If closed-end responses are used on the questionnaire, computerised scanning can be used for data entry. It has been estimated that in the Women's Health Initiative study of 160,000 women, the use of a machine-read FFQ cost US$1.2 million, compared to an estimated US$23.2 million for 3-day diet records, or US$25 million for three 24-hour recalls (18).

A disadvantage of the FFQ is the time and cost involved in developing the questionnaire. Creating a food list, which is the perfect balance between including enough foods, and not including too many, can be a time intensive task. A long food list can create participant fatigue. On the other hand, a shorter list may exclude important sources of nutrients, reducing the accuracy of the questionnaire. The questionnaire also needs to be kept up-to-

date with current food trends and new products on the market. Each updated version then needs to be validated in a test population to ensure the quality is still at an acceptable level (7).

2.2.2 Challenges of Dietary Assessment in Toddlers

Random and systematic errors are a part of all dietary assessment methods (15). The quality of the data, however, will vary depending on the population and techniques and care taken during the data collection process. It is important to understand the challenges involved with measuring the diet of a toddler so that errors can be minimised, and data can be interpreted appropriately.

The most obvious challenge when measuring a toddler's dietary intake is the need to rely on a surrogate reporter. A toddler does not have the cognitive ability to correctly identify foods eaten, let alone accurately recall foods eaten in the past (19). Studies that have investigated the accuracy of a child's mother reporting on behalf of their child, found only small differences between the mother's report and an observers report, from meals eaten either the same day (20) or the previous day (21). This however is a simplistic view, as the mother (or primary caregiver) is not always present to feed every meal to their child. Toddlers often attend day-care centres, are looked after by another family member, and can have multiple carers throughout a day (6). Baranowski et al (1991) found that when a group of 3-5 year old children's mothers were away from their child for greater than 4.5 hours per day, 15% could provide no useful information, and 37% could only provide information on dietary intake for part of the day during a 24-hour recall (22). However when looking only at mothers who could provide information for the full day, not-at-home mothers and at-home mothers reported dietary intake with the same accuracy. It is possible that the inability to collect information from not-at-home mothers is reduced when assessing diet with an FFQ instead of the 24-hour recall. This may be because an FFQ asks about average usual dietary intake over a period of time as opposed to exact dietary intake the previous day. Results of previous research have shown that participants find it easier to describe usual frequency as opposed to describing a specific meal in the past (23). Participants often estimate frequencies by either determining an approximate rate, and extrapolating this out to the required reference period or by taking a 'usual' estimate as the starting point, and then increasing or decreasing this depending on whether they thought the reference period was a 'usual' period or not (23). This may mean that the effect of the primary caregiver being

absent for some meals is diluted by the fact they can use other 'clues' as to the frequency of the foods eaten. This is reinforced by Parrish et al (2003) who found it did not affect the validity of an FFQ whether or not the parent was the sole meal provider for the child (14).

Another issue unique to measuring the diet of toddlers is the amount of leftovers. Toddlers typically do not eat all of their meal, and the uneaten food is not always left on the plate. Food can be thrown away, spilt on their body, or fed to pets, creating a harder task for the caregiver to estimate how much of the original portion was actually eaten. A study by Dubois (1987) found that even dieticians could not accurately estimate leftovers, generally overestimating the portion eaten, and underestimating the amount left over (24). It would be expected that a dietitian would be more aware of portion size estimates than the general public, and all food was left on the plate for estimation in this assessment. This demonstrates the difficult task caregivers have when estimating the portion eaten by their child.

2.3 Components of a Food Frequency Questionnaire

A new FFQ does not necessarily need to be created every time one is required for use, modification of an already existing questionnaire may be more appropriate. Careful consideration needs to be undertaken of the original purpose of the FFQ, the characteristics of the population it was designed for, how acceptable the level of validity was, and when it was developed (17). In a review of 227 validation studies, 45% were found to be adapted versions of previous FFQs, with no difference in FFQ validity between the original and adapted questionnaires (25). If there is no suitable FFQ available for use in the population to be studied, a new FFQ will need to be developed. It is important to consider the purpose of the FFQ during the development process to ensure it is designed to reflect the usual intake of the nutrients of interest.

2.3.1 The Food list

The questionnaire should be designed to ask the minimum number of questions which will provide the optimal amount of required information. Extra questions increase the length of the questionnaire and can result in participants losing focus and making mistakes (15). Food items included on the food list need to be commonly eaten by a number of participants, and/or contain a substantial amount of the nutrients of interest. They also need to be discriminating, and therefore be eaten in varying amounts by the participants (17). For

example, if every participant ate one carrot every day, this would be a good source of carotene, but would not provide useful ranking information. Spinach however, with a lower carotene content, which is eaten commonly by some participants, and not at all by others, would be more informative for the ranking of participants (7).

The number of foods listed in FFQs ranged from 5 to 350, with a median of 79, in the review article by Cade et al (2002). The length of the questionnaire was mainly determined by whether FFQs were designed to assess single nutrients or to obtain information on many nutrients. Willett et al (1998) suggests that a food list of 100 is close to the maximum number of questions a participant can complete before fatigue and boredom begins to reduce the quality of the answers (7).

The order of the food list is also important. Food groups that are the most important should be placed near the beginning, but not right at the start. Participants are more likely to make errors in the first couple of questions as they work out how to answer the questionnaire, and then again towards the end, as they tire (26). Related foods should be grouped together with the more specific foods placed before the more general questions, such as 'coated biscuits' followed by 'other biscuits' (7). Multiple clear questions, asking about one food at a time, are easier to understand and interpret than a longer more complex question asking about multiple foods (17,27). When more than one food is asked about in a question, the subject needs to adjust the frequency and portion size to account for both foods, increasing the time taken to answer the question and this can become confusing for the participant. Although asking about a single food in each question will make the questionnaire longer, it may actually reduce the overall time taken to complete the questionnaire (7,27).

2.3.2 Frequency Response Options

Frequency responses can be obtained using multiple-choice or open-ended questions. An open-ended question, where the answer is given in terms of frequency per day, week, or month, can allow for more information to be collected than multiple-choice options. However, it has been found the longer coding time, increased transcription errors, and higher non-completion rate of open-ended questions means that multiple-choice options are preferred (7,17,27).

Multiple-choice frequency questions usually have five to ten options available (7). The purpose of an FFQ is to capture the between-subject variation, and therefore foods with a

high nutrient content eaten rarely (e.g. liver), and foods eaten often (e.g. bread), need to be captured. Therefore frequency options usually range from, 'never' through to 'over six times per day', however 'two or more per day' at the higher end, has been deemed sufficient for non-beverage items (7,27). Foods eaten commonly will usually provide most of the information on nutrient intake, therefore while there should be no gaps in the frequency options, particular emphasis on the higher weekly frequencies is important (15,17).

2.3.3 Portion Size Estimation

Information on usual portion size is not always collected. A questionnaire which asks only about the frequency of consumption is called a qualitative FFQ (8). Qualitative FFQs provide no information on nutrient intake, and only classify participants with respect to the frequency of foods eaten. At a later stage, average portion sizes may be allocated to each food eaten by the participants (17). However, this reduces the sensitivity of the FFQ as it will not be able to discriminate between two participants who eat foods with the same frequency but in different amounts (28).

A quantitative FFQ assesses both the frequency and the amount of each food eaten. Multiple choice questions can be used which allow the participant to describe the portion as 'small', 'medium' or 'large', where the size of the medium portion is usually stated. These terms however, are subject to widely varying interpretations (29). Another option is to use an open ended question where a description of the portion size is given in relation to a measurement aid. The most commonly used aids are household measurements, drawings or photographs, and food models (8). Another option would be to use the child's palm volume. A child's palm is generally closer in size to the portions eaten by a toddler than other measurements such as cups and bowls, and therefore possibly easier for a caregiver to use when estimating portion size. There is, however, no literature currently available on the validity of its use. Measurement aids in general, however, have been shown to produce fewer errors compared with unaided responses, and even fewer errors when a range of sizes are shown and not just one average size displayed (15,28,30,31). It appears however, that no one measurement aide is better than another, with the largest measurement error believed to be associated with quantifying portion size (8,29), suggesting further research into measurement aids is required.

Another way to include the portion size in the questionnaire is to use a semi-quantitative approach, which asks how often a certain portion size is eaten (8). By incorporating the portion size into the frequency question, the number of questions is reduced, however not necessarily the time taken to complete the survey. Combining the two questions into one can be confusing for some participants particularly if they do not eat the food in the allocated portion size. It is expected that the subject adjusts the frequency to account for the different portion size, but in reality, it is unlikely that this always happens, with many participants simply ignoring the stated portion size (7).

Some authors have found that frequency accounts for the greatest between-person variation and therefore argue collecting portion size data may have no added value (32,33). This conclusion however is counter-intuitive, as in theory, more information should provide more accurate results. It is possible that when portion size is included, the improvement in accuracy from its addition is masked by the errors involved in estimating the portions. This suggests that instead of removing portion size from FFQs, ways to increase the accuracy of the participants' estimates needs to be explored instead (15).

2.4 Designing a Validation Study

Validation studies are undertaken to examine the relationship between the best estimate of the truth and what was measured (15). Using an unvalidated method may result in inaccurate relationships between diet and disease being found. Epidemiological studies often require a fast easily applied approach to dietary assessment, but this needs to be balanced against getting accurate answers. The accuracy of the method is determined by undertaking a validation study where the 'test' method is compared with another method, the 'reference' method (8). Ideally the validity of a dietary method is the degree of closeness between the true value and the value given by the test method (i.e. accuracy) (7,8). The reference method will not be completely valid and therefore the relationship between the two methods will provide information on the 'relative validity' of the method. The relative validity cannot indicate if the FFQ is accurate or not, only whether or not the answers are comparable to the reference method (17).

It is also important to assess the reproducibility (or reliability) of the FFQ, which is a measure of how well the FFQ can reproduce the same results multiple times (i.e. precision) (17). There are two types of reproducibility which can be assessed: agreement between FFQs

collected by different interviewers with the same participant (interrater reliability) and agreement between FFQs collected by the same interviewer and participant (intrarater reliability) (34).

The degree of reproducibility for the two measurements limits the amount of agreement possible between the two different methods of dietary assessment (35). It is also possible for an FFQ to have high reproducibility but low validity, and so it is important for both reproducibility and validity to be assessed (7).

When designing a validation study, there are many factors which need to be taken into account to ensure the results are a true representation of what would be collected when the questionnaire is used in real life, such as the population the participants will be recruited from.

2.4.1 Study Population

The FFQ needs to be validated in a sample of the population which it is intended to be used in. Factors such as age, ethnic group, sex, and health status of a population can affect how questions are answered and therefore the study results (17). Perceptions about healthy and unhealthy foods differ between cultures, and may influence the reported levels of consumption for some foods (15). Socioeconomic status and ethnicity can also affect the validity due to varying diversity in the diet. A diet of low diversity increases agreement, as there are fewer opportunities available for mistakes to be made, compared to one of high diversity. The questionnaire's food list is also likely to be aimed at a particular group of society and may be missing some foods that are eaten often in another culture, reducing its appropriateness.

2.4.2 Number of Participants and Replicate Days Required

The sample size required is influenced by many factors, including the statistical methods used to assess reproducibility and validity, the precision required, the daily variation of the nutrients of interest, the number of days recorded by the reference method, and the time frame the days are recorded over (7,36). Within- and between-subject variation is also different across populations (i.e. different countries, ethnicities, sex, age) therefore making it very difficult to accurately estimate the number of participants required.

Cade et al (2002) suggests that Bland and Altman statistics requires at least 50 participants, and preferably more than 100 (17). Correlation coefficients require approximately 150 participants before there is no longer an increase in precision of the confidence intervals (CI) after this number (7). These sample size numbers however, presume greater than 12 days of food intake recorded (7,17), which is not feasible in most validation studies.

These recommendations are also made for adults, and are likely to be higher than what is required when investigating a toddler population. This is because the number of days required to estimate usual dietary intake is influenced by the variance ratio (within- to between-person variation). The variance ratio is significantly lower in children less than four years old compared with adults (37). It has been shown that children adjust their energy intake at meals so that their within-subject daily energy intake is relatively stable. Adult validation studies also tend to focus on FFQs that investigate dietary intake over longer time periods such as a year (7,25). Dietary intake over a year is likely to have a higher variation than intake over the past month, due to factors such as seasonal availability of certain foods.

It is generally agreed that no more than 100-200 participants are required, with a sufficient number of days recorded to accurately describe usual diet (7,17,38–40). A study conducted by Lanigan et al (2004) in toddlers up to 2 years old, found that to achieve a hypothetical correlation coefficient of 0.9, at most, five days of weighed diet record was required (41). Aligned with this, Stram et al (1995) and Willett (1998) found that the ideal number of days is five in most situations. Collecting more than five days starts to become inefficient due to participant burden (7,42). This can result in fewer participants completing the study, a higher potential for food consumption to be altered, and the increasing awareness of foods eaten may affect the participants' answers to the FFQ (7).

2.4.3 Reference Method

To be able to assess validity, the FFQ must be compared to a reference method. In choosing a reference method, the sources of error must be largely independent of the errors associated with an FFQ (17). The main errors of an FFQ are its reliance on portion size estimation, recall of foods eaten, and an incomplete list of foods. The weighed diet record is the optimal method for comparison as it has errors least likely to be correlated with an FFQ. When completing a diet record, participants weigh all the foods and drinks at the time of consumption, and therefore memory, portion size estimation and a restricted list of foods

are not errors associated with a diet record. In practise however, portion size estimation is still an issue as there are often times when the participant cannot weigh the foods. A diet record also has high participant burden and often a dietary recall is used instead. These can be useful in less educated populations, in particular with participants with limited literacy (17). They are also less likely to influence the actual diet of the subject, as the subject often does not know when they will be undertaken. However, because the recall requires reliance upon memory and portion size estimation, its errors are not as independent from the FFQ as those of a diet record. Another option is to use biomarkers as the reference method. They have few errors associated with an FFQ, as they are independent of dietary intake (8). However, they are nutrient-specific and only some nutrients such as protein and vitamin C are able to be assessed (8). Biomarkers are therefore usually used alongside another reference method as opposed to being used as an alternative. Moreover, they also contain errors, as factors such as digestion, absorption, metabolism and excretion all affect the relationship between the biochemical marker and usual dietary intake (17). They also increase participant burden and are invasive, and therefore are not as appropriate for a toddler population as an adult population.

The period over which the days the diet record is recorded on should cover the same period the FFQ asks about. For example if the FFQ asks about the previous four weeks of intake, the diet record days should be evenly spread over four weeks (17). Ideally, the days should be non-consecutive, and spread evenly across each day of the week. Consecutive days have been shown to be correlated, and therefore non-consecutive days provide a better example of the true variability of the subject's usual diet (36,43).

2.4.4 Sequence of Administration

The sequence of administering the reference method and the FFQ is important as one may influence the other. Participants would normally encounter the FFQ independent of the reference method and the validation process should mimic this. The effort required in recording all foods consumed for the diet record could increase participants' awareness, and therefore artificially increase their accuracy in completing the questionnaire if the FFQ was administered afterwards. On the other hand, administering the questionnaire before the diet record could result in artificially low associations, as the questionnaire would relate to the dietary intake before the period assessed during the diet record.

It is also important to consider the time over which both methods will be undertaken. Willett et al (1998) suggests it is appropriate to administer the FFQ before and after the test method (7). The results can then either be averaged, or one of the two FFQs for each participant randomly selected and compared with the test method. This reduces the disadvantages associated with using solely the FFQ administered before or after the test method, and allows for the reproducibility of the questionnaire to be assessed (15).

The time between the two administrations of the FFQ must not be too close, as reproducibility results may be affected by participants remembering previous answers. Validity may also be affected by the 'learning effect', as participants may be more aware of how to answer the questions on the second attempt. On the other hand, the uniqueness of undertaking a study in a toddler population means a shorter time frame is preferable. Only one multi-nutrient FFQ has been assessed for reproducibility in toddlers, and it used a four week period between the two administrations (12). Food habits are rapidly changing throughout childhood as food preferences become more developed and evolved (44). If the time between administrations of the dietary assessment methods is too long, inconsistencies in the results between administrations may be due to real dietary changes, rather than poor performance of the FFQ.

2.4.5 Nutrient Database

The one source of error common to both the FFQ and the diet record is the food composition database used. If possible, it is desirable to use the same database for both dietary assessment methods so as to match any errors, and reduce their effect on the assessment of validity.

When choosing a database to use, it is advisable to use a national database which is comprised of foods relevant to that population and with nutrients appropriate for the foods found locally. Using a database which is regularly updated is important as foods change in composition regularly, new foods are developed, and preparation methods change over time (15). The number of foods and nutrients included in a database vary and the completeness and accuracy of the nutrients needs to be considered when choosing which one to use (36). Too many substitutions of foods will decrease the accuracy of the diet record.

2.4.6 Statistical Analysis

2.4.6.1 Validity

There are a range of statistical techniques available to assess validity by comparing two dietary methods. There is however, a lack of consensus on which are the most appropriate (45).

2.4.6.1.1 Comparison of Means

Comparing nutrient means or medians for the data obtained using the two methods assesses the relative validity at a group level (8). A paired t-test or Wilcoxon's signed rank test, depending on the distribution of the data, can be used. (46). Such tests are useful when absolute intakes are important, and where differences between participant groups are required (17). It must be remembered though, that a comparison of group means provides no information on the quality of the questionnaire at the individual level, or its ability to describe the distribution of intakes (47).

2.4.6.1.2 Correlation Coefficients

Correlation coefficients are the most commonly used statistics in dietary assessment validation studies (48). Pearson correlations are used for normally distributed data and Spearman rank correlations if the data cannot be normalised by log transformation (8). The Spearman rank correlation can also be used when the objective of the study is to rank participants, and may in fact be preferable to Pearson correlations as it is less sensitive to extreme values due to its rank order approach (45).

If there is large within-subject variation in nutrient intakes, the correlation between the two methods will be decreased. This affect can be accounted for by calculating the ratio of the within- to between-subject variation (the variance ratio) and using the number of days recorded by the reference method to calculate an attenuation factor. This can then be used to adjust the correlation and 95% confidence intervals (CI) calculated (8).

A review of FFQs found that validity was usually accepted if nutrient correlations were above 0.4 (25). This is in line with Willett et al (1998) which considers correlations of 0.30-0.50 to be 'acceptable', and 0.50-0.70 to be 'good' (7).

The use of correlations as a measure of validity however, is controversial. Bland and Altman (1986) argue that as the two methods are measuring the same thing, there will always be a

positive correlation, whereas the null hypothesis for the test is that there is no expected correlation (35). They also argue that correlations only show association and not agreement, and poor agreement can exist when there is high correlation between the two methods (49). High agreement will only exist when the two methods produce very similar results, whereas high correlation can still occur if the test method is a fixed proportion higher or lower than the reference method (35).

Correlations are also affected by the characteristics of the study population. If there is high between-subject variation the correlation will be higher than for a group with a lower between-subject variation (7,48). The range of values in the sample, which in turn is partially affected by the sample size also affects the correlation (17); the smaller the sample, the higher the correlation (50).

It is for these reasons correlation coefficients have been deemed by some to be too flawed to assess validity (17,35,51). However, when used in conjunction with other statistics they can be useful (17,45). As they are so commonly used in validation studies, it is also useful to report them to enable comparisons to be made between studies (25).

2.4.6.1.3 Classification into Categories of Consumption

Cross-classification involves classifying participants into categories, usually tertiles, quartiles or quintiles, of nutrient intake based on both the test and reference method (8). The percentage of participants who are classified into the same categories (correctly classified) and opposite categories (grossly misclassified) by the test and reference methods can be calculated. This percentage agreement will however include a percentage of participants classified by chance alone (7).

The Kappa statistic can be used to take into account the percentage correctly classified as well as the proportion of participants expected to be correctly classified by chance alone (52). When looking at ordinal variables it is advised to use the weighted Kappa, which gives greater emphasis to larger differences (i.e. gross misclassification) than to smaller differences (i.e. adjacent cells). However, the Kappa statistic is limited in that it depends on the number of categories used, and the weightings allocated to each category. A greater number of categories provide a higher potential for disagreement and therefore a lower Kappa value possibly understating agreement (34). Caution should therefore be used when comparing different studies measured with different categories or weightings.

2.4.6.1.4 The Bland-Altman Method

Bland and Altman (1999) have strongly advocated the use of the limits of agreement in method comparison studies (51). This involves plotting for each nutrient the difference between the two methods versus the average of the two methods, and calculating the limits of agreement and their corresponding 95% CI. This plot allows outliers to be seen easily, as well as whether any trend is present with increasing intakes.

A larger difference between the reference method and the FFQ is often found when participants have a high nutrient intake (judged by the average of the FFQ and diet record), than in participants with a lower nutrient intake (17). This means that the FFQ will appear more reliable in participants with a lower nutrient intake than in those with a higher nutrient intake.

Bland and Altman (1986) argue that this technique is the best way to look at the agreement between two methods, and it has been recommended for use in dietary assessment by an international group of experts (25,35). The method is however assessing the validity of the FFQ at the individual level, and so its appropriateness depends on the overall purpose of the questionnaire.

2.4.6.2 Reproducibility

Reproducibility can be affected by factors inherent in the questionnaire. If the questionnaire contains pre-determined portion sizes, or contains restricted frequency options, less variability is permitted and therefore reproducibility is increased. Real dietary change will also affect the results if the time between questionnaires is too long (17,47). As for validity, there is little consensus on the most appropriate statistics to use for assessing reproducibility. The choice of statistic is guided by the purpose of the FFQ with the statistics subject to the same limitations discussed in section 2.4.6.1.

To assess the FFQ at the group level, comparison of means (paired t-test) or medians (Wilcoxon matched-pairs signed-rank test) can be used to see if they are statistically different (8). At the individual level, cross-classification and the weighted Kappa can be calculated to assess how well the FFQ separates participants into classes. The Bland-Altman method and limits of agreement can be used to assess the agreement between the two methods, also at the individual level (8).

Correlations are not considered to be an acceptable method to assess reproducibility as they do not assess agreement, only association, which would be expected as you are comparing repetitions of the same questionnaire with the same participant (25). It may, however, be useful to report correlations so that findings can be compared with other studies, providing more appropriate methods are also used (17).

Use of the Pearson correlation in adult reproducibility studies is decreasing due to its inability to detect systematic error (53), with the use of intraclass correlation coefficients (ICC) becoming more common (53,54). Intraclass correlations take into account the degree of correlation and extent of disagreement between methods including correcting for any agreement which occurred by chance (8,46). Intraclass correlations are sensitive to systematic bias (54), and are not influenced by sample size (55). However, the ICC, as with the other correlation coefficients, is sensitive to the sample heterogeneity and therefore results can only be interpreted for a population similar to the tested sample.

2.5 Currently Validated Food Frequency Questionnaires for use in Toddlers 12 to 24 months old

Few FFQs have been validated for use in preschool children and only seven appear to have been designed for use in toddlers aged 12-24 months (9–14,56). As one of these studies only investigated food groups, not nutrients, it has been excluded from this review (56). The study design and findings from the six studies validated for multi-nutrients have been summarised in **Table 2.1**.

The results from these studies suggest that FFQs generally overestimate energy and nutrient intakes compared to the reference method (9–14). In three out of the six studies, all nutrients and energy intake were overestimated (9,13,14) and for the other three studies, 11 out of 16 (10), 13 out of 18 (11), and 14 out of 19 (12) nutrients were overestimated. All but one study overestimated energy, with the percentage difference between the FFQ and reference method ranging from 7 (10) to 70% (14).

It was generally commented by the authors that portion size estimation was the main area responsible for the overestimation of intakes. Four studies used household measurements and natural portion sizes such as ‘slices’ of bread (9,10,12,13). Although these are common portion sizes which the caregiver or parent should be able to visualise, it may be too difficult

to scale down these sizes to one appropriate for a toddler typically eating much smaller portions. One study allocated average portion sizes from national data (11), however this approach is limited in that it does not allow questionnaires to be sensitive to any variation in portion size. It has also been shown in adult studies that allowing participants to describe their portion size increases the validity (25).

The differences in overestimation by the studies may also be explained in part by the quality of the questionnaire and the validation study (39). The time period and sequence of administration of the FFQ and reference method is important as answers to the FFQ can be influenced by the reference method if administered last, but when administered first the FFQ will be covering a different time period to the reference method (7,8,17). One study averaged the results of FFQs taken before and after the reference method (11), while the other studies administered the FFQ once, either before (9,10,12,13) or after (14) the reference method. An additional option would be to randomly use either the participants first or second FFQ data for comparison with the reference method. This would provide a combination of the minimal and maximal validities, and therefore reduce any bias introduced by using just the first or second FFQ (7).

Macronutrients tended to be less overestimated (up to 100% difference) compared to micronutrients (up to 250% difference) (9–14). This has also been found in adult studies, and is thought to be a result of many foods each contributing a small amount to total micronutrient intake, as opposed to a small number of foods contributing the majority of total macronutrient intake (39). Correlations between the FFQ and reference method also varied. Some correlations were weak e.g. 0.18 for vitamin D (9), and 0.25 for vitamin E (13) while other studies found high correlations: 0.63 for magnesium (11) and 0.74 for calcium (12). However some nutrients such as fibre, vitamin A and vitamin B_{12} were relatively consistent across studies. Correlation coefficients for fibre varied between 0.23-0.38 (9–12); vitamin A 0.34-0.50 (9–11), and vitamin B_{12} 0.35-0.47 (11–13). In contrast, other nutrient correlations showed greater variation across studies, such as vitamin C 0.19-0.58 (9–14), and calcium 0.26-0.65 (9–13). This variability between nutrients may be due in part to the high within-person variation, that is, the true day-to-day variation of dietary intake, particularly when the FFQs are looking at intake over different lengths of time (7). There will be higher variation of dietary intake over a year, such as in the Parrish et al (2003) study (11), compared with only 14 days in the Blum et al (1999) study (14). It is generally conceded

however, that the FFQs provided comparable results to those validated in adult populations (correlations usually 0.4 to 0.7 (7)) and therefore are an appropriate tool for assessing toddlers' diets at the population level.

There are many factors which influence the validity of dietary assessment methods, some of which are unique to toddlers. Two studies compared results for children who were fed solely by their parents versus the results of children with multiple carers (10,14). They found no differences in FFQ validity between these two groups, even though the FFQ was only completed by a parent and the reference method by all carers. Two studies looked at whether changing food patterns during the study period affected correlation coefficients. Marriot et al (2008) found that during the study 22% of participants changed from baby milk formula to drinking cow's milk. When all foods and drinks containing milk were excluded from the data, the range of nutrient correlation coefficients increased from 0.25-0.66 to 0.44-0.70 (13). This demonstrates that the change in the type of milk in the toddler's diets, due to the differing nutrient compositions, had a significant effect on the correlations. This change in diet had a significant effect on the validity as the FFQ was completed before the diet record, and therefore any changes were only recorded in the diet record and not the FFQ. In contrast, Andersen et al (2003) found that although 67% of participants said they had begun introducing new foods to their toddler during the study period, when the data were split into two groups based on whether the caregiver had said their child's diet had changed or not, the introduction of new foods did not significantly affect the validity of the FFQ (9). Andersen et al (2003) also administered the FFQ before the diet record days, however it is possible that no effect was found due to foods being introduced that were more similar in nutrient content than the different types of milk in the study by Marriot et al (2008).

When looking at the diet of toddlers, quantifying breast milk is an important consideration. One study used an algorithm based on suckling time and published intake data to estimate breastmilk intake (13). In contrast Parrish et al (2003) decided to leave out all breast milk data from both the FFQ and reference method, reasoning that the available breastmilk intake data were too unreliable (14). The other studies did not state whether they included breastmilk or not (9,11,12). Between 4% and 58% of toddlers were breastfed during the study period in these studies. If breastmilk was excluded from the analysis a large section of these toddlers' diets would be missed, therefore reducing the ability of the FFQ to reflect the total diet. Nevertheless, often the only option is to leave it out of analysis, as reliable

published data on breastmilk intake for the population of interest is not readily available. The other option is to calculate data by recording the amount of breastmilk consumed and its nutrient content; however this requires test weighing and breast milk sample collection and analysis, a difficult and invasive process.

Finding an FFQ which is suitable for New Zealand toddlers is difficult as food patterns change over time and so a questionnaire suitable for measuring intake of New Zealand toddlers also needs to have been validated recently, to ensure it is up-to-date with current food trends and available products (17). One of the FFQs in this review was validated three years ago and another four years ago and so both are relatively recent (12,13). The rest were validated 8-13 years ago and would need to be updated and revalidated to ensure their appropriateness for current use.

FFQs must also be validated in a sample of the population they are going to be used in, as different backgrounds can influence the way in which participants respond to the questions (8). An FFQ which has validated well in a highly educated population may have instructions or questions which are confusing and difficult to answer for a less-educated population. If the questionnaire is validated in the population of interest then these mistakes would be picked up and changes could be made to the questionnaire to make it more appropriate. Food choices also differ between countries and countries also have foods that are unique to them, such as kumara in New Zealand, which is commonly eaten and therefore important to be included in an FFQ for New Zealanders. Two of the FFQs were validated in low-income populations (11,12) and none were validated in New Zealand. These FFQs are only suitable for use in the population they were validated in so there is a need for an FFQ to created or modified, and then validated for use in New Zealand toddlers. If the decision was made to modify and validate an existing FFQ it would make sense to start with one that performed well in a robust validation study in the population it was originally used in.

The sample sizes used in almost all the validation studies were low, with only two out of the six studies having more than 100 participants (10,11). It is recommended that validation studies have 100-200 participants, depending on the number of replicate measures of the reference method and the accuracy required (17), as discussed in section 2.4.2.

Three studies used the 24-hour recall as the reference method (11,12,14). The 24-hour recall is not the most suitable method to validate an FFQ against, as both rely on participant

memory and portion size estimation. The errors from the two methods need to be independent otherwise the agreement may be falsely increased (7). As it is not possible in dietary assessment to ever know the absolute 'truth', a method of higher quality with unrelated errors is recommended. The weighed diet record, unlike the 24-hour recall and FFQ, allows for direct measurement of portion sizes and does not rely on memory as foods are recorded at the time of eating and it is therefore the most appropriate choice (7) as used in three of the studies (9,10,13). It is also recommended that the reference method is recorded on non-consecutive days due to the correlation between foods eaten on consecutive days (36,57). Only three studies (11,12,14) assessed dietary intake on non-consecutive days.

One study used an FFQ which asked about the previous year of intake (14). Although the participants were recruited when they were 12 months old, by the time they did the FFQ a year later they would have been at least 24 months or older. It is likely that this questionnaire is not suitable for use when the child is between 12 and 24 months old, particularly if the child is at the lower end of this age range. It is also quite difficult to ask a participant to accurately recall the usual food intake of their child between the ages of one and two years. The food habits of a toddler are rapidly developing between these ages as at one year they are still being introduced to new foods, and by the time they are two years old most are eating the same foods as the rest of their family.

It can be concluded from this review that more studies are needed regarding the validity of the FFQ for use in toddlers 12-24 months old before they can be accepted as a valid tool. A questionnaire can be created or modified from an existing FFQ, but will still need to be validated in a robust study before it can be used.

Table 2.1: Validity of food frequency questionnaires for assessing nutrient intake in toddlers 12 to 24 months old

Reference	Population Characteristics	FFQ	Reference Method	Study Design	Results
Andersen (2003) (9)	64 Norwegian toddlers 12 months old 39% attended childcare 58% breastfed	FFQ designed to describe dietary habits at 12 months and feeding practices birth to 12 months. Author does not describe development of FFQ Asks about food intake on previous 14 days 44 questions: 4 about breastmilk, 3 about other milk, 18 about 140 food items grouped together, 3 about dietary supplements and 17 about other food habits. Frequency alternatives ranged from never/less than once a month to several times per day, author did not mention total number of categories Portion size estimated from photographic booklet with a series of four different portion sizes for each food item. Also household units used for some foods	7-day weighed diet record: Four consecutive days ↓ One week interval ↓ Three consecutive days	FFQ completed ↓ One week interval ↓ 7-day diet record begun	FFQ overestimated all nutrients except calcium, and energy intake by 25% compared with diet record Correlation coefficients ranged from 0.18 for vitamin D to 0.72 for polyunsaturated fatty acid (median r=0.50) Spearman rank correlation coefficients (absolute intake): Energy 0.43 Protein 0.57 Total fat 0.56 Carbohydrate 0.25 Fibre 0.23 Vitamin C 0.41 Vitamin B_{12} - Calcium 0.62 Iron 0.62 Zinc - Average correctly classified into quartiles was 36%, and 4% misclassified into extreme quartiles

Table 2.1 continued

Reference	Population Characteristics	FFQ	Reference Method	Study Design	Results
Andersen (2004) (10)	187 Norwegian toddlers 24 months old 75% attended childcare or had a nanny 6.4% breastfed	FFQ designed to describe dietary habits at a group level and to rank individual intakes. FFQ based on: previous dietary surveys in toddlers, general knowledge about this age group and experience from adolescent and adult nation-wide surveys Asks about food intake on previous 14 days 39 questions: 4 on milk, 15 about 125 food items grouped together, 3 about dietary supplements and 17 on other food habits Frequency alternatives ranged from never/less than once per month to several times a day, author did not mention total number of categories Portion size estimated from photographic booklet with four different portion sizes for each food Also household units used for foods not included	7-consecutive day weighed diet record	FFQ completed ↓ 2-week interval ↓ 7-day weighed diet record	FFQ overestimated 11 out of 17 nutrients and energy intake by 7% compared with the diet recall Correlation coefficients ranged from 0.26 for total fat and iron to 0.50 for vitamin A (median r=0.38 for absolute intake) Spearman's Correlation Coefficients: Energy 0.31 Protein 0.27 Total fat 0.26 Carbohydrate 0.38 Fibre 0.38 Vitamin C 0.39 Vitamin B_{12} - Calcium 0.26 Iron 0.42 Zinc - Average correctly classified into quartiles was 36%, and 5% misclassified into extreme quartiles

Table 2.1 continued

Reference	Population Characteristics	FFQ	Reference Method	Study Design	Results
Blum (1999) (11)	239 low-income children from North Dakota, USA 1 to 5 years old, 55% 1 to 2 years old Author did not state percentage of children attending day care or percentage breastfed	Adapted from the Harvard Service FFQ designed to assess the diets of low-income adult women Asks about food intake over the past 4 weeks 103 questions: 84 on food, 19 on food habits, supplements and services Frequency options ranged from never to six or more times a day, with 9 response options Portion sizes were derived from age appropriate national data	Three 24-hour recalls, two on weekdays and one on a weekend day Recalls completed over a month, approximately every 7 to 10 days Author did not state how portion sizes were estimated	FFQ completed ↓ Three 24-hour recalls over the next month ↓ FFQ completed again	FFQ overestimated 11 out of 19 nutrients compared with the diet recall, energy by 0.24% Correlation coefficients ranged from 0.26 for dietary fibre to 0.63 for magnesium, (average r=0.52 adjusted for energy intake (EI) and within-person variation) Pearson's Correlation Coefficients (adjusted for EI and within-person variation: Energy - Protein 0.43 Total fat 0.62 Carbohydrate 0.52 Fibre 0.26 Vitamin C 0.58 Vitamin B_{12} 0.47 Calcium 0.60 Iron 0.51 Zinc 0.31 Author did not assess cross-classification

Table 2.1 continued

Reference	Population Characteristics	FFQ	Reference Method	Study Design	Results
D'Ambroio (2012) (12)	22 low-German-Speaking Mennonities from Mexico, living in Canada, aged 12 to 36 months old Low-income mothers Author did not state % attending day-care Author did not state % breastfed	Purpose of the FFQ to assess mean nutrient intakes FFQ adapted from the Children's Nutrition Questionnaire (CNQ) (11) (which is adapted from the Harvard Service FFQ). Adapted through focus groups The CNQ asks about the previous 4-weeks of dietary intake, however author did not state time period of the adapted version 97 food questions Author did not explain the frequency response section Portion sizes estimated using measuring utensils and food models	Three 24-hour, multiple-pass recalls (n=14) Eight participants only completed the first 24-hour recall Author did not mention how portion sizes were estimated	FFQ and 24-hour recall completed on day 1 (n=22) ↓ 2 weeks after baseline, 24-hour recall (n=14) ↓ 4 weeks after baseline, 24-hour recall and FFQ (n=14)	FFQ overestimated 5 out of 11 nutrients and energy by 9.6%, compared with the diet recall Correlation coefficients ranged from 0.19 for vitamin C to 0.80 for caffeine (average r=0.41) Spearman rank correlation coefficients: Energy 0.34 Protein 0.52 Total fat 0.25 Carbohydrate 0.30 Fibre 0.38 Vitamin C 0.19 Vitamin B_{12} 0.35 Calcium 0.74 Iron 0.30 Zinc - Author did not assess cross-classification

Table 2.1 continued

Reference	Population Characteristics	FFQ	Reference Method	Study Design	Results
Marriott (2008) (13)	50 UK infants aged 12 months Author did not state % of participants attending childcare 14% were breastfed	FFQ developed to describe dietary patterns and rank individuals according to nutrient intake Developed using a review of data from a UK survey of 18 month old children, 24-hour recalls of mothers attending a baby clinic, and food diaries of 12-month-old preterm infants Asked about previous 28 days Contained 78 food questions Frequency options ranged from never to an open ended question for multiple times a day, with 10 response options Prompt cards were used to show the foods included in each group Portion size estimated using household measures and food models	Four day weighed diet record 72% of records were consecutive days	FFQ completed ↓ Four diet record days (median interval between the two methods seven days)	FFQ overestimated all nutrients and energy by 19%, compared with the diet record Correlation coefficients ranged from 0.25 for vitamin E to 0.66 for sodium (median r=0.49) Spearman rank correlation coefficients: Energy 0.46 Protein 0.52 Total fat 0.36 Carbohydrate 0.49 Fibre - Vitamin C 0.52 Vitamin B_{12} 0.42 Calcium 0.65 Iron 0.44 Zinc 0.39 After excluding milk from the analysis the range of correlation coefficients increased to 0.44 (fat) to 0.70 (protein and sodium) Author did not assess cross-classification

Table 2.1 continued

Reference	Population Characteristics	FFQ	Reference Method	Study Design	Results
Parrish (2003) (14)	68 Canadian children 1-3 years old 51% of children were fed meals on a regular basis by a caregiver other than their parents 4% were breastfed	Used the Willet FFQ 111 questions Asked about usual intake over previous year No mention of frequency options No mention of how portion size was estimated	Four 24-hour recalls, one approximately every three months No mention of how portion size was estimated	Four 24-hour recalls conducted over 1 year ↓ FFQ completed	FFQ overestimated all nutrients and energy by 70%, compared with the 24-hour recall Correlation coefficients ranged from 0.08 for energy to 0.42 for vitamin C (mean r=0.32) (nutrient correlations adjusted for energy) Pearson correlation coefficients: Energy 0.08 Protein 0.33 Total fat 0.39 Carbohydrate 0.41 Sugar - Fibre - Vitamin C 0.42 Vitamin B_{12} - Calcium - Iron - Zinc - Author did not assess cross-classification

3 Methods

This study was designed to validate an FFQ to capture habitual intake of toddlers aged 12 to 24 months over the preceding four weeks. This FFQ is designed to be completed by the primary caregiver.

3.1 Food Frequency Questionnaire Development

3.1.1 Modification

The original FFQ developed for use in this study was an extended and revised version of the semi-quantitative questionnaire used in the Southampton Women's Survey (SWS), designed to assess nutrient intake and dietary patterns in 12-month old infants (13). The SWS questionnaire was interviewer-led, contained 78 questions, and assessed the previous four weeks of dietary intake. The modifications of the SWS FFQ were undertaken by another student for use in the Prevention of Overweight in Infancy (POI) study (Appendix A and B) and so was not part of this candidate's remit. However, the process will be discussed in detail below to enable clarification of the food choices within our 'EAT' FFQ.

3.1.2 The Food List

The food list of the questionnaire was modified to include foods consumed on a daily basis by at least 10% of New Zealand toddlers, as reported in previous research (58). As this data is 13 years old, supermarket tours were also undertaken in Dunedin to identify any new products likely to be commonly eaten by New Zealand toddlers. Foods were included on the basis of the products making at least a similar contribution to shelf space (indicating sales volume) to foods already on our list. Toddler and infant foods were also identified on these tours to ensure they were products available in New Zealand.

Several food sections were altered, with the addition of 'dairy and dairy products', and the removal of the 'desserts', 'spreads' and 'miscellaneous' sections. The foods contained in these sections were either incorporated into other sections (e.g. custard moved into the 'Dairy and dairy products' section) or removed altogether (e.g. peanut butter). Spreads were removed, as although they are a food eaten frequently by toddlers, the small amounts eaten provide very few nutrients to their diet. Cross-check questions were also added to the

'vegetable' and 'fruit' sections, which ask for the total number of vegetables/fruits eaten. This allows an adjustment factor to be calculated so that the frequency of each individual item within the group can be altered, a technique shown previously to reduce the overestimation prone to occur when a long itemised list is included in an FFQ (59).

Some questions that included multiple foods were separated into individual questions. For example the SWS FFQ had 'breakfast cereals and porridge' as one question. This was separated into four questions: 1) 'weet-bix, fruity-bix etc', '2) porridge', '3) cornflakes, rice bubbles', and '4) other breakfast cereals'. Other combined questions were left together, such as 'meat pies and sausage rolls'. Given the unknown variability of the nutrient content in complete meals, items such as 'meat casseroles, stews, and curries' were removed. Instead, it was expected that participants would include each ingredient individually under the appropriate question; for example carrots in the vegetable section and beef in the meat section.

3.1.3 The Frequency-response Options

The number of frequency options remained unchanged from the original SWS FFQ, however more instructions on how to complete the questionnaire were provided. For example 'never' was changed to 'Tick box if not offered this month', as the original wording may have been confusing for the participant had their child not eaten the food in the past four weeks, but had eaten the food before this.

3.1.4 Portion Size Estimation

The original questionnaire asked for the average amount per serving in terms of household portion sizes or food models. For example, cheese was asked in terms of; '1 tbsp grated = 0.5, 1 slice (see drawing) = 1, 1 tbsp cottage = 1, small triangle = 1'. It was thought that having multiple options was confusing for the participant, and therefore simplifying the questionnaire to include only two different types of units (one for each food) would make the questionnaire more user-friendly.

All portion estimates were modified to be answered in natural portion sizes such as 'number of kiwifruit' or 'number of crackers'. Any foods that did not have a natural portion size were asked in terms of the child's palm volume (e.g. the 'number of palms' of rice eaten). This is defined as the length, width, and thickness of the child's palm, as shown in Figure 3.1 (refer to section 3.2.3.3 for further information). Palm volume was chosen to be used as a child's

palm will always be in close proximity to the food they are eating, and therefore possibly easier for the caregiver to recall when completing the questionnaire. A child's palm is also relatively small, and therefore closer to the size of the actual servings eaten than other measurements such as cups and bowls.



Figure 3.1: Palm Volume; surface area of the palm and thickness

Another novel aspect of this questionnaire was that it also took into account any leftovers by asking for portion size information using two questions; the average amount offered each time, and then the amount actually eaten. This is an important differentiation for the caregivers to make, as with toddlers, there can be a lot of plate wastage, and the amount eaten is what is important, not the amount offered.

3.1.5 Additional Questions

Additional questions in the SWS FFQ were kept, asking for the number of meals per week given to the toddler by someone other than the primary caregiver completing the questionnaire. If this answer was greater than zero, how many of these meals were able to be included in the questionnaire ('none', 'some', 'most' or 'all') was asked.

The SWS FFQ had a detailed section at the beginning of the questionnaire asking about the child's breastfeeding history. This section was removed for the EAT FFQ, and one line each added for infant formula, toddler milk and breastmilk to the 'dairy and dairy products' section. The frequency of the child's consumption of breastmilk was asked, but information on the amount of breast milk could not be ascertained. Dietary supplements were excluded from the EAT questionnaire as it was intended that the information would be collected in a separate questionnaire, so the validation focused only on dietary sources of nutrients.

3.1.6 Expert Review and Pre-testing

The FFQ was reviewed by a group of experts and pre-tested in a group of approximately 20 toddlers 12-24 months old. Some minor changes were made as a result, such as adding '1 baby rice cake = 3 rice wheels' to the key in the 'rice cake' food line, resulting in the finalized EAT FFQ.

3.1.7 Finalised EAT Food Frequency Questionnaire

The EAT FFQ was designed to be quantitative, so that it could assess nutrient intake and describe dietary patterns in toddlers 12 to 24 months old.

The questionnaire included questions on eleven food groups (baby/toddler food; bread & crackers; breakfast cereals; rice & pasta; meat, chicken, fish, eggs & beans; vegetables; fruit; dairy & dairy products; cakes, biscuits & snacks; drinks; other foods & drinks) using a total of 98 food and drink related questions.

3.1.8 Coding of the Food Frequency Questionnaire

Nutrient lines for the FFQ were calculated using FOODfiles 2010, except for toddler and baby foods, by this candidate (60). Nutrients for toddler and baby foods were calculated by contacting the manufacturer (in the case of infant formulas and toddler milks), or by looking at the nutrient label and food list on the product container. Recipes were then made in Kai-culator (Department of Human Nutrition, University of Otago, New Zealand, 2011) in line with the recipe list and nutrient labels.

When multiple foods were collapsed into one food line on the questionnaire, the foods were weighted according to the reported frequency of consumption and average portion size eaten, using data reported previously (58). For example, previous data showed 61% of toddlers ate potato with an average serving size of 64g, and 10% of toddlers ate kumara, with an average serving size of 25g. The FFQ item 'potato and kumara' was then weighted 6% kumara and 94% potato on the basis of the following calculation for potato: proportion eating potato x average potato serving size, divided by the sum of the proportion eating potato x average potato serving size + proportion eating kumara x average kumara serving size (61%*64g)/((61%*64)+(10%*25)). The five different ways to prepare potato were then equally divided to make up the 94%, and the three ways to prepare kumara were divided equally to make up the 6%.

Participants reported portion sizes in units of volume, and therefore a density of each food item was required to enable conversion to a gram weight. The density of each FFQ item was taken from those present in FOODfiles 2010, and where foods were missing, from the previous version, FOODfiles 2006 (60,61). When an item in the questionnaire was made up of more than one food, densities were weighted according to the percentage that food contributed to the total nutrient line. For example, the FFQ item 'rice' was made up of 95% white rice (density=0.74) and 5% brown rice (density=0.82). The density of the FFQ item was then 0.74, on the basis of the following calculation: proportion of white rice x density of white rice + proportion of brown rice x density of brown rice ((95%*0.74) + (5%*0.82) = 0.74).

3.2 The Validation Study

3.2.1 Ethical Approval

The study was conducted by researchers from the Department of Human Nutrition, University of Otago. The Human Ethics Committee of the University of Otago, Dunedin, granted ethical approval for the study (Appendix C). Māori consultation was undertaken (Appendix D), and written informed consent was obtained from all primary caregivers at the first appointment.

3.2.2 Recruitment and Study Population

A required sample size of 200 was estimated, based on recommendations in the Cade et al (2002) review paper, and the requirements of another MSc student using the data (17). The review article recommended between 100 and 200 participants should be recruited for a validation study. The other MSc student using the data intended to conduct principal component analysis on the data which requires 10 participants per food group. It was therefore decided to recruit 200 participants so that 20 food groups could be analysed.

Participants were recruited from September 2011 through to April 2012, in Dunedin, Mid-Canterbury and Wellington, following the study advertising protocol (Appendix E). Study advertisements were displayed in local newspapers, posters (Appendix F), flyers, emails and on internet web pages as described in detail in **Table 3.1**. All participants received a nutrient analysis of their child's diet (Appendix G), a healthy eating pamphlet produced by the

Ministry of Health New Zealand (62,63), and were entered into one of two draws to receive an iPod Touch.

Participants were toddlers aged between 12 and 24 months and their primary caregivers, who were living in either Dunedin, Mid-Canterbury, or Wellington. Primary caregivers were eligible to take part in the study if:

- they were responsible for care of a child aged between 12 and 24 months
- the child was born at > 36 weeks gestation
- the child had no diagnosed illness known to affect growth and/or food consumption

Table3.1: Recruitment activities in Dunedin, Mid-Canterbury, and Wellington

	Dunedin	Mid-Canterbury	Wellington
Recruitment Period	September – December 2011	December 2011	September 2011 – April 2012
Newspapers	The Star • Three advertisements • Free delivery to 43,500 homes	The Snowfed • One advertisement with article alongside • Free delivery to 1,600 homes The Rakaia News • One advertisement • Free delivery to 1,200 homes The Ashburton Guardian • One article promoting the study • Delivered to 5,500 paid subscribers	The Wellingtonian • Five advertisements • Free delivery to 70,000 homes Northern Courier and Petone Herald • One advertisement in each • Combined free delivery to 70,000 homes
Posters	40 posters displayed in: 12 Plunket rooms, Moana swimming pool, toddler movement and music groups, supermarkets, dairies, Medical Practices, the Parents Centre, and community notice boards	15 posters displayed at: Ashburton Swimming Pool, cafes, supermarkets and day-care centres in Ashburton, Methven and Rakaia	20 posters displayed at: Te Papa, Wellington School of Medicine, community noticeboards, cafes and day-care centres
Flyers	80 flyers distributed to: The Parents Centre Dunedin, child clothing stores, and Day-Care Centres	• 120 flyers distributed with the Ashburton Parents Centre newsletter • 100 flyers placed in children's lunchboxes at day-care centres • 80 flyers handed out at Music and Movement classes	10 flyers handed out at 'Baby Rock and Rhyme' in Karori Library, Wellington

Table 3.1 continued

	Dunedin	Mid-Canterbury	Wellington
Emails	37 emails to day-care centres (emails were forwarded on and/or advertisements printed in their newsletters) One email sent to all University of Otago staff and postgraduate students based at the Dunedin campus		12 emails sent to day-care centres and coffee groups (emails forwarded on to parents and/or advertisements printed in their newsletters) Two emails sent to all University of Otago staff and postgraduate students based at the Wellington campus
Internet webpages	'Conscious Parenting Group Dunedin' Facebook page	Ashburton Trading Society web page	'Breastmates' and 'Wellington North Parents Centre' Facebook page Te Papa Intranet

Once primary caregivers had signified their interest they received a phone call from a researcher during which their eligibility was checked (Appendix H). If participants were eligible their first study appointment was booked. The primary caregiver was then sent an information letter confirming their appointment time, an information sheet and consent form (Appendix I).

3.2.3 Data Collection

The study comprised of two visits following the appropriate protocols (Appendix J and K). Dunedin and Mid-Canterbury participants were visited at home, whereas Wellington participants either came to the Otago University School of Medicine (Wellington), or if this was not possible, visited at their home. At the first visit primary caregivers signed the consent form, completed a brief background questionnaire including questions such as date of birth and ethnicity, the child's weight, length and palm size was measured, the FFQ was completed for the first time, and caregivers received the diet record along with scales and instructions. Over the next four weeks, the diet record was completed on five randomly allocated days, and once completed, a second visit was conducted. At the second visit the diet record was returned and checked by the researcher for any mistakes. Primary caregivers also completed the FFQ for a second time. More details on each aspect of data collection follow in sections 3.2.3.1-3.2.3.4.

All researchers involved with data collection received training on how to administer the FFQ, provide instructions on completing the diet record, and the collection of length, weight and palm measurements, according to standard operating protocols.

3.2.3.1 Background characteristics

At the first visit, primary caregivers were asked to complete a questionnaire containing socio-demographic questions relating to themselves and their child, as well as their child's birth weight, birth length and length of gestation (Appendix L). Socio-demographic questions included: primary caregiver's age, relationship to child in the study, number of children born to the caregiver, and the caregiver's and child's ethnicities. Questions relating to ethnicity were taken from the New Zealand 2007 census (64)

3.2.3.2 Anthropometrics

All measurements relate to the child and were taken at the first visit by trained researchers, following the anthropometric protocol (Appendix M) adapted from the World Health

Organization (WHO) guidelines (65). All equipment was calibrated daily, the scales to within 0.1kg and the rollameter to within 0.1cm.

3.2.3.2.1 Child's Length

The child's length was measured using a Rollameter (Harlow Healthcare Rollameter 100, UK). The Rollameter was placed on a hard flat surface and covered with a thin cloth for hygiene purposes. If the child was wearing any hair ornaments, the caregiver was asked to remove these, and lay the child on his/her back against the fixed headboard, compressing the hair. The head was positioned so that the crown touched the head board and the Frankfort Plane was perpendicular to the board. The caregiver was asked to hold the child's head in this position.

The child was to lie straight along the board, shoulders touching the board, spine not arched and arms resting against the sides of the trunk. The researcher then moved the footboard against the child's feet, while holding the child's legs down with one hand to ensure they were straight. If both legs could not be held in this position, the measurement was taken with one leg in position. The soles of the feet were flat against the footboard, toes pointing up. Duplicate measures were taken to the nearest 0.1cm. If the duplicate measures differed by greater than 0.7cm, a third measurement was obtained, and the two closest measurements used to create the average.

3.2.3.2.2 Child's Weight

Weight was measured with digital scales (Seca Alpha Model 770; Seca, Hamburg, Germany) placed on a flat, hard, even surface. The child was weighed wearing a nappy (diaper) of known weight (30 grams) supplied by the researcher. They were also able to leave a singlet on if one was worn. If the child was capable, they could stand on the scales by themselves, otherwise the caregiver was asked to stand on the scales, the scales tared, and then the child given to the caregiver to hold and the weight recorded. Duplicate measures (accurate to the nearest 0.1kg) were taken. If the measurements differed by greater than 0.1kg, a third measurement was taken, and the two closest measurements averaged.

3.2.3.3 FFQ Data Collection

The FFQ was completed twice, at the first and second visits approximately four weeks apart. The caregiver present completed the questionnaire with assistance from the researcher, where necessary. When answering the questions, the caregiver was asked to keep in mind

their child's diet over the past four weeks. For more information on the contents of the questionnaire see section 3.1.1.

3.2.3.3.1 Palm size

Palm thickness was measured only at the first visit using an anthropometer (Model 01291). Duplicate measurements were required to be within 0.2 cm, otherwise a third measurement was taken and the two closest measurements averaged (Appendix M). Thickness was measured in the middle of the child's palm; usually the deepest part of the palm, with the palm facing upwards so that the base plate of the anthropometer was on the back of the palm. Accuracy of the measurements was taken to the nearest 0.1cm.

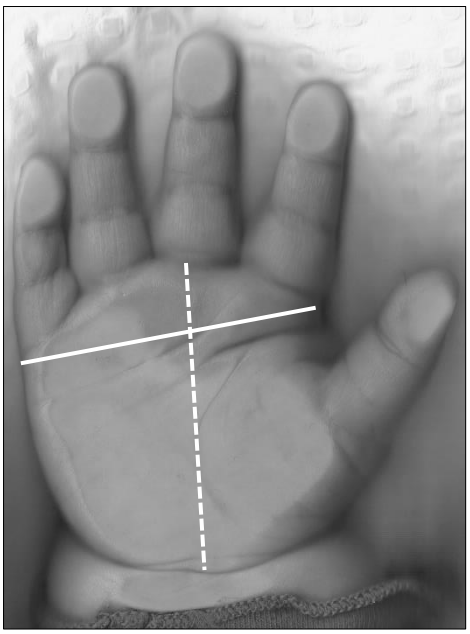


Figure 3.2: Scanned image of a child's hand, with length and width measurements

The child's palm was scanned using a portable scanner (Canoscan LiDE 110; Cannon, USA) to ascertain the length and width. The right hand of the child was placed on the scanner surface, with the hand naturally opened, wrist extended and the third finger close to the measuring tape. The length and width of the palm were then measured using the computer program ImageJ (Free Software Foundation, Boston, USA). Length was taken to be the distance from the middle of the wrist crease to the middle of the base of the third digit. Width was the maximum width across the metacarpal – phalangeal joints II and V (see Figure 3.2). Palm volume was then able to be calculated by multiplying the child's palm length, width and thickness to give the portion volume.

If a palm scan was unable to be taken for a child, their palm length and width was calculated by averaging the palm length and width of all other participants who were within two months of their age and also had a palm thickness within ± 0.2cm of their measured palm thickness.

3.2.3.4 Diet record

After completing the FFQ, all participants received instructions during their first visit on how to complete a five-day weighed diet record, and were provided with a food diary (Appendix N), a set of dietary scales (Salter Electronic Model Selectronic 2200) and two spare batteries. Scales were accurate to within ± one gram. The food diary contained written instructions on the first four pages, and an additional page contained written instructions on how to use the scales and change the batteries if necessary (Appendix O).

The toddler's weighed diet record was completed by their primary caregiver for five days over a four week period. The first day of recording was the day immediately following the first interview, with the next four days randomly allocated. Days were allocated so that there was an approximately even number of every day of the week collected from the total sample. All participants received a reminder text the day before each of their specific diet recording days (Appendix P), and a phone call after their first day to check for any errors and questions they may have had.

On each allocated day, participants were asked to weigh and record all foods and drinks consumed, including water. They were asked to record the time and location of where the food was eaten and the name, brand and cooking method used to prepare the food. The participant was then asked to weigh and record the weight of the plate or mug, and then weigh each food item, and record the weight, without taring the scales between foods. They were then asked to weigh all the leftovers and attempt to estimate the amount of each food making up the leftovers (e.g. all of the potato, 80% of the broccoli and a teaspoon of pasta sauce). At a later date the researcher then calculated the 'amount eaten' by subtracting the leftovers from the amount offered.

The diet record contained a question asking the parent if the child was unwell on the day of recording, and if this influenced their child's appetite by increasing or decreasing it.

Participants were also supplied with a supplementary page containing photographs of commonly eaten 'takeaway foods' (e.g. McDonald's fries, pizza) (Appendix Q). Each photo contained a food item with a corresponding weight, and participants were able to estimate from the photos the approximate weight of the food their child was offered. The supplementary page also had a ruler and set of circles for measuring any food items that could not be weighed.

The caregiver present at the meeting was asked to instruct any other carers on how to record foods and drinks eaten by their child at times when they were not present. An instruction sheet with examples and a blank page similar to the diet record was also provided to give to other carers (Appendix R). If the primary caregiver provided the food while their child was with another carer, the primary caregiver was asked to weigh the food items beforehand, and request all leftovers to be returned so they could be weighed.

At the second appointment when the diet records were returned, all items recorded on the diet record were checked by a nutritionist to identify any mistakes, missing foods, or illegible handwriting. These were then clarified by the primary caregiver.

3.2.4 Data Entry and Nutrient Calculations

Data from the demographic questionnaire and anthropometry measurements were entered into an Excel spreadsheet (2007 Microsoft Office Corporation).

3.2.4.1 Diet Records

The diet records were entered into Kai-culator (Department of Human Nutrition, University of Otago, New Zealand, 2011), which uses the New Zealand Food Composition Database (New Zealand, FOODfiles 2010) for all foods except toddler and infant foods to calculate intakes of energy and nutrients (60). The toddler and infant milks use manually entered data obtained from manufacturers. The toddler and infant foods used data from the previous New Zealand Food Composition Database (New Zealand, FOODfiles 2006) (61).

In the diet record, participants provided all ingredients where possible for homemade recipes, and either the proportion of the total recipe offered to the child, or the gram amount offered. Raw ingredient weights were converted to cooked weights using conversion factors provided by Food Standards Australia New Zealand (66). An Excel spreadsheet was

then used to calculate percentages of each ingredient so they could be entered based on the proportion of the total cooked recipe given to the child as recorded in the diet record.

All diet-records were checked twice for accuracy and consistency of data entry. The first check ensured all food items and amounts were correct. The second check corrected any items entered using the standard Kai-culator recipe instead of individual ingredients (e.g beef lasagne). 'Unlikely' foods such as raw meat were checked, and corrected where appropriate.

3.2.4.2 Food Frequency Questionnaire

The FFQ data were entered into LimeSurvey version 1.91+ (Free Software Foundation, Germany) and were entered twice, each time by different trained researchers using the data entry protocol (Appendix S). Any data entry errors were identified and reconciled.

The number of grams eaten per day of each food item from the FFQ was then calculated by multiplying the frequency per day, volume (in grams), and the amount eaten. Volume was either in palms or natural portions. If the volume was in palms, volume was calculated by multiplying the length, width and thickness of the child's palm, by the density of the food item. Nutrients were then able to be calculated from the grams eaten per day of each food item by dividing by 100, and multiplying by the nutrient per 100g. For information on the nutrient line development, see section 3.1.8.

The EAT FFQ contained 13 questions for vegetables and 11 questions for fruit. A greater number of items within the same category can over-estimate actual intake, and therefore cross-check questions were included in these two sections. These asked the caregiver the total number of fruits or vegetables their child ate each day, week or month. The FFQ was originally designed with one question in each section: 'How often has your toddler had fruit/vegetables in the past month', however it was found in the Prevention of Overweight in Infancy (POI) study, that caregivers were answering this by recalling the number of meals which contained fruit or vegetables, and not the total number of servings eaten. The FFQ was then modified for this validation study to include a second part to the question: 'When you think about one of these occasions how many different fruits/vegetables would you offer'. It was expected that participants would average the number of fruits/vegetables eaten at each time counted in the first question, to answer this question.

This meant that for each participant, for every fruit and vegetable, the unadjusted data was multiplied by an adjustment factor. For POI adjusted fruit intakes, the adjustment factor was the total frequency of fruits eaten per day (e.g. the participants answer to the first question), multiplied by the sum of all the individual frequencies in the fruit section. The EAT adjusted fruit intakes adjustment factor, was the total frequency of fruits per day (e.g. the participants answer to the first question) multiplied by the number of fruits usually offered per serving (e.g. the participants answer to the second question), divided by the sum of all the individual frequencies in the fruit section. This was calculated again following the same method for the vegetable section in the questionnaire. These adjustments allowed us to determine the reproducibility and validity of the FFQ as it stands (unadjusted), but also when adjusted for fruit and vegetable intake as used in the POI study (67) and in this study.

3.2.5 Statistical Analysis

All statistical analyses were undertaken on Stata/IC statistical software version 12.0, 2012 (StataCorp, College Station, TX, USA). Continuous data from the descriptive statistics were checked for normality using the Shapiro-Wilk test, and visually by producing histograms with a normal curve superimposed. The data was shown by both tests to be skewed, so geometric means and 95% confidence intervals (CI) were calculated for continuous variables, and numbers and percentages for categorical variables. Descriptive statistics were calculated separately for participants (those who completed the diet record and at least one of the two FFQs) and non-participants.

Relative validity of the FFQ for assessing energy and nutrient intakes was calculated by randomly using either the first or second FFQ from each participant and comparing this with the diet record data. This was done by randomly generating a list of numbers (either '1' or '2') using an Excel spreadsheet and allocating these numbers in order of appearance to the study identification codes. If a '1' had been allocated to a participant their first FFQ was used, and if a '2' was allocated, the second FFQ was used. Nutrients of interest were: total fat, carbohydrate, protein, fibre, calcium, iron, zinc, vitamin B_{12}, vitamin C, and energy.

The following statistics were undertaken with the unadjusted FFQ data, and then repeated for POI adjusted, and then EAT adjusted FFQ data.

The mean nutrient intake data for both the diet record and FFQ was assessed for normality using the Shapiro-Wilk test and visually using a histogram with a normal curve superimposed

on the graph. Geometric means and 95% CIs were calculated for both the FFQ and the diet record. A comparison of means was calculated using a paired t-test with the level of significance set at a p-value of less than 0.05. Spearman's rank correlation coefficients were used to assess the association between the FFQ and diet record data for each nutrient.

The nutrient intakes were then divided into quartiles relating to the absolute intake for each nutrient according to both the FFQ and diet record. It was decided not to report the Kappa statistic as the Kappa value depends on the number of categories and weightings used, and therefore cannot easily be compared between studies (34). Instead, the proportion of diets that were classified by the FFQ to the same quartile as the diet record and to opposite quartiles, was then calculated to assess the ability of the FFQ to categorise the child's diets (5,6). Bland-Altman plots were used to assess the agreement between the two methods at the individual level, with mean percentage agreement and limits of agreement, expressed as percentages. The limits of agreement were calculated as mean difference ± 1.96 standard deviations of the differences (7).

The reproducibility of the questionnaire was assessed by comparing the first and second administrations of the FFQ. The mean daily energy and nutrient intake data for both administrations of the FFQ was assessed for normality using the Shapiro-Wilk test, and visually using a histogram with a normal curve, superimposed on the graph. A comparison of means was calculated using a paired t-test with the level of significance set at a p-value of less than 0.05. Intraclass correlation coefficients were calculated to assess the association between the first and second FFQ for energy and each nutrient (5,6).

4 Results

4.1 Recruitment and Participants

A total of 160 eligible participants were recruited at baseline, of which 152 completed both FFQs and the diet record, with one additional participant completing just the first FFQ and the diet record. This meant that 153 participants (96%) could be included in the validity analysis and 152 participants (95%) in the reproducibility component. A palm scan and therefore palm length and width measurements could not be obtained from two of these participants (palm thickness was obtained). Length and width estimates for these participants were obtained by averaging the palm size of all other participants who were within two months of their age who also had a palm thickness within ± 0.2cm of their measured palm thickness.

The participants who withdrew from the study tended to be more likely to have a child who was female, less likely to be of 'New Zealand European or other' (NZEO) ethnicity, had only one child, and were from homes with low deprivation (NZDep2006 1-3) **(Table 4.1).** The 153 primary caregivers, who completed an FFQ and the diet record for their child, had an average age of 34.4 years. Almost all (95.6%) were the child's mother, with 4.4% being the child's father. The children had an average age of 16.8 months, 51.3% were male, and had a mean length-for-age percentile of 58 (95% CI: 52,65), and mean weight-for-age percentile of 67 (95% CI: 62,73) using the New Zealand growth charts (68), which are based on the UK growth charts, which in turn are based on the WHO research of six reference populations with optimal growth (69). Caregivers described their child as sick on a total of 18% of the diet record days. On these sick days, 59% of children were thought to have a decreased appetite whereas 4% had an increased appetite, and 37% no change.

Table 4.1: Characteristics of the toddlers who completed the study and those who withdrew

Characteristics	Study Participants[a] (n=153) *n (%)* or mean (95% CI)[c]	Participants who withdrew[b] (n=7) *n (%)* or mean (95% CI)
Sex		
Male	80 (52.3)	2 (28.6)
Female	73 (47.7)	5 (71.4)
Age (months)	16.8 (16.2,17.4)	16.4 (14.0,19.1)
Ethnicity[d]		
NZEO	137 (89.5)	5 (71.4)
Māori	12 (7.8)	2 (28.6)
Pacific Islanders	4 (2.6)	0
Birth Characteristics		
Birth weight (kg)[e]	35.5 (34.5, 36.4)	39.0 (34.8,43.6)
Birth length (cm)[f]	51.9 (51.1,52.8)	50.0[f]
Length of gestation (weeks)[g]	39.6 (39.4, 39.9)	40.5 (39.9, 41.1)
Baseline Measurements		
Weight (kg)	11.0 (10.8, 11.2)	10.5 (9.0, 12.3)
Length (cm)[h]	81.7 (80.9, 82.5)	77.3 (72.8, 82.1)
Siblings		
Only child	86 (56.2)	5 (71.4)
One sibling	48 (31.4)	2 (28.6)
Two or more siblings	19 (12.5)	0
NZDep2006 Index of Deprivation[ij]		
1-3	60 (42.0)	4 (80)
4-7	67 (46.9)	1 (20)
8-10	16 (11.2)	0
Breastfed[k]	44 (28.8)	na[m]
Supplement user[l]	14 (9.2)	na[m]

[a] 'Study participants' are those who completed at least one FFQ and the diet record (one participant completed the diet record and the first FFQ only).

[b] 'Participants who withdrew' are those who signed the consent form but did not complete the diet record and therefore their data were excluded from statistical analysis.

[c] Geometric mean and 95% confidence interval (CI).

[d] Ethnicity was prioritised in the order: Māori, Pacific Island, then New Zealand European and Others (NZEO).

Data were only available for [e]150 study participants and 6 withdrawn participants, [f]74 study participants and 1 withdrawn participant (therefore no CIs were available), [g]151 study participants and 5 withdrawn participants, [h]153 study participants and 6 withdrawn participants.

[i] The NZDep2006 Index of Deprivation divides New Zealand geographically into deciles of deprivation. The least deprived 10% of New Zealand is given an index of 1, conversely, the 10% most deprived population in New Zealand is given an index of 10 (70).

[j] Data were only available for 143 study participants and 5 excluded participants.

Children recorded as having [k] any breast milk or [l] supplements during the diet record.

[m] Not available

4.2 Relative Validity

Tables 4.2-4.7 contain data presented as unadjusted and adjusted dietary intakes. The FFQ contained a two-part cross-check question at the beginning of both the fruit and the vegetable section. The cross-check question contained two parts: a) 'How often has your toddler had vegetables/fruit in the past month?' and b) 'When you think about one of these occasions how many different vegetables/fruits would you offer?'. The 'POI' adjusted results use dietary intake adjusted only by part a, and the 'EAT' adjusted results have taken into account both part a and part b.

4.2.1 Energy and Nutrient Intakes

Table 4.2 presents the geometric mean intakes of energy and selected nutrients measured by the FFQ and 5-day diet record (5DDR) as unadjusted and adjusted values. The FFQ is a combination of the first and second FFQs, selected by randomly choosing one FFQ from each participant. The unadjusted FFQ data gave significantly higher estimates of energy and nutrient intakes compared to the diet record for all nutrients (p-value < 0.05). Similar results were obtained when the data were adjusted by either technique, with the exception of POI adjusted vitamin C which was not significantly different from the 5DDR (p-value $= 0.096$).

4.2.2 Correlation Coefficients

Spearman correlation coefficients between the FFQ and the 5DDR are presented in **Table 4.3**. Unadjusted correlations averaged 0.50, and ranged from 0.37 (zinc) to 0.66 (calcium). The correlations generally increased, albeit by a small amount, when the fruit and vegetable sections were adjusted by the cross-check questions. The mean POI adjusted correlation was 0.52 (range: 0.36 zinc-0.68 calcium). EAT adjusted data did not provide a further

Table 4.2: Average daily intakes of energy and selected nutrients in the 5DDR and FFQ

	5DDR	Unadjusted FFQ[bc]	POI adjusted FFQ[bc]	EAT adjusted FFQ[bc]
Nutrient	Mean (95% CI)[a]	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)
Energy (kJ)	3494 (3347, 3647)	5251 (4994, 5521)	4646 (4421, 5081)	5317 (5053, 5595)
Protein (g)	33 (32, 35)	54 (51, 57)	51 (49, 54)	55 (52, 58)
Total fat (g)	28 (27, 30)	49 (46, 52)	45 (43, 48)	50 (47, 53)
Carbohydrate (g)	111 (107, 116)	149 (141, 157)	124 (118, 131)	151 (143, 159)
Fibre (g)	7.9 (7.5, 8.4)	13.4 (12.6, 14.3)	9.6 (9.0, 10.2)	13.8 (12.9, 14.8)
Calcium (mg)	556 (511, 605)	820 (758, 888)	782 (720, 850)	830 (769, 897)
Iron (mg)	5.8 (5.4, 6.2)	7.4 (6.9, 7.9)	6.5 (6.0, 6.9)	7.5 (7.0, 8.0)
Zinc (mg)	4.2 (4.0, 4.4)	7.2 (6.8, 7.6)	6.7 (6.4, 7.1)	7.3 (6.9, 7.7)
Vitamin B_{12} (mg)	1.5 (1.4, 1.7)	2.4 (2.3, 2.6)	2.4 (2.3, 2.6)	2.4 (2.3, 2.6)
Vitamin C (mg)	47.5 (42.8, 52.8)	75.4 (69.1, 82.4)	36.6 (40.2, 49.0)	77.3 (70.0, 85.5)

[a] Geometric mean and 95% confidence interval (CI)

[b] The FFQ result is a combination of the first and second FFQ, where one FFQ was randomly selected from each participant.

[c] A two-part cross-check question was included at the beginning of both the fruit section and the vegetable section of the FFQ. 'Unadjusted FFQ' presents the energy and nutrient results before the cross-check questions are used to adjust energy and nutrient intakes. 'POI adjusted FFQ' is adjusted by only the first section of the cross-check question, and 'EAT adjusted FFQ' results are adjusted by both parts of the cross-check question.

Table 4.3: Spearman correlation coefficients between the food frequency questionnaire and the 5-day diet record, compared with results from previous FFQ validation studies in toddlers

Energy/nutrient	Results from previous studies[a]	Unadjusted[b]	POI adjusted[b]	EAT adjusted[b]
Energy (kJ)	0.08-0.46[f]	0.48	0.50	0.53
Protein (g)	0.27-0.57[g]	0.48	0.49	0.50
Total fat (g)	0.25-0.62[g]	0.37	0.40	0.42
Carbohydrate (g)	0.25-0.52[g]	0.58	0.59	0.55
Fibre (g)	0.23-0.38[e]	0.59	0.56	0.57
Calcium (mg)	0.26-0.74[f]	0.66	0.68	0.67
Iron (mg)	0.31-0.48[c]	0.61	0.63	0.61
Zinc (mg)	0.30-0.62[f]	0.37	0.36	0.40
Vitamin B_{12} (mg)	0.24-0.47[d]	0.41	0.39	0.41
Vitamin C (mg)	0.19-0.58[g]	0.48	0.57	0.51

[a] The range of correlations from six multi-nutrient FFQ validation studies in toddlers 12-24 months old (9–14). A correlation coefficient was only available from [c] two studies, [d] three studies, [e] four studies, [f] five studies, or [g] six studies. Three studies reported Spearman's correlations (9,10,13), two studies reported Pearson correlations (12,14), and one study reported energy-adjusted Pearson correlations (11).

[b] A two-part cross-check question was included at the beginning of both the fruit section and the vegetable section of the FFQ. 'Unadjusted' presents the Spearman correlation coefficient before the cross-check questions are used to adjust energy and nutrient intakes in the FFQ. 'POI adjusted' results are for the FFQ adjusted only by the first section of the cross-check questions, and 'EAT adjusted' results are for the FFQ adjusted by both parts of the cross-check questions.

increase in the mean correlation compared to POI adjusted intakes (mean: 0.52; range: 0.40 zinc-0.67 calcium). Correlations were comparable with the other studies, generally being at the higher end of what has been previously observed, with energy, carbohydrate, fibre, and iron correlations all higher, irrespective of any adjustment.

4.2.3 Bland-Altman Method

Unadjusted mean percentage agreements between the FFQ and 5DDR were all greater than 100% (range: 128% for iron – 172% for total fat), showing the FFQ overestimated mean energy and nutrient intakes compared to the 5DDR **(Table 4.4)**. POI adjusted vitamin C had a mean percentage agreement of 93% (95% CI 85,103), which indicates good agreement between the FFQ and the 5DDR as the 95% CI includes 100%. Energy and all other nutrients were overestimated by the POI adjusted FFQ data (range: 112% carbohydrate and iron – 160% zinc) and EAT adjusted FFQ data (range: 130% iron – 175% total fat) compared to the 5DDR. EAT adjustment did not substantially alter the discrepancy. POI adjustment however, decreased the discrepancy between the FFQ and 5DDR for some nutrients such as iron which decreased from 128% when unadjusted to 112%.

The difference between the upper and lower limits of agreement varied for each nutrient, such as unadjusted carbohydrate (limits of agreement: 79%-227%) compared to unadjusted vitamin C (limits of agreement: 46%-551%). POI and EAT adjusted intakes had narrower limits of agreement than the unadjusted intakes, however EAT adjusted limits of agreement were wider than the POI adjusted intakes, except for vitamin B_{12} which was not changed substantially by any level of adjustment.

Table 4.4: Bland-Altman statistics measuring strength of agreement for energy and nutrient intakes between the FFQ and 5DDR[a]

Nutrient	Unadjusted[b]		POI adjusted[b]		EAT adjusted[b]	
	Mean % agreement (CI)[c]	Limits of Agreement (%)	Mean % agreement (CI)[c]	Limits of Agreement (%)	Mean % agreement (CI)[c]	Limits of Agreement (%)
Energy (kJ)	150 (144,157)	89-254	133 (128,138)	81-219	152 (146,159)	91-256
Protein (g)	164 (156,172)	89-303	155 (148,163)	84-285	166 (158,174)	90-304
Total fat (g)	172 (164,182)	88-337	159 (151,168)	83-308	175 (166,184)	91-334
Carbohydrate (g)	134 (128,140)	79-227	112 (107,116)	67-186	135 (130,141)	78-235
Fibre (g)	169 (159,179)	82-348	121 (114,128)	59-248	174 (164,185)	81-374
Calcium (mg)	148 (140,155)	79-227	141 (134,148)	76-262	149 (142,157)	80-279
Iron (mg)	128 (120,136)	58-280	112 (106,119)	52-242	130 (122,139)	59-286
Zinc (mg)	170 (162,179)	89-325	160 (152,168)	83-306	173 (164,182)	91-326
Vitamin B_{12} (mg)	159 (148,171)	65-388	159 (148,171)	65-390	159 (148,171)	65-389
Vitamin C (mg)	159 (144,175)	46-551	93 (85,103)	28-311	163 (147,180)	46-580

[a] Data from the 5DDR and FFQ was natural log-transformed, back-transformed and multiplied by 100%.

[b] A two-part cross-check question was included at the beginning of both the fruit section and the vegetable section of the FFQ. 'Unadjusted' presents the energy and nutrient results before the cross-check questions are used to adjust energy and nutrient intakes in the FFQ. 'POI adjusted' results are for the FFQ adjusted only by the first part of the cross-check questions, and 'EAT adjusted' results are for the FFQ adjusted by both parts of the cross-check questions.

[c] Mean % agreement = FFQ/5DDR (%), CI = 95% confidence interval of the mean % agreement

4.2.4 Cross-classification

The percentage correctly classified ranged from 30.1% (total fat) to 48.4% (carbohydrate) for unadjusted data. Adjustment made little difference to the findings; results remained consistent, with approximately one third to one half correctly classified. The average percentage correctly classified did not vary from the unadjusted value of 40.1% to an EAT adjusted value of 40.9% **(Table 4.5).** The average percentage gross misclassification also remained relatively unchanged, ranging from 2.9% (EAT adjusted) to 4.0% (unadjusted data) with the differing levels of adjustment.

4.3 Reproducibility

4.3.1 Energy and Nutrient Intakes

Energy and nutrient intakes are displayed in **Table 4.6** as unadjusted and adjusted values for FFQ-1 and FFQ-2. FFQ-1 refers to the questionnaire conducted during the first interview at baseline, and FFQ-2 refers to the questionnaire conducted at the second interview, on average, 5.0 weeks later (range: 3.6 to 8.7 weeks). POI adjusted intakes tended to be lower than the unadjusted values for both FFQs, and EAT adjusted intakes tended to be higher than the unadjusted values. There were no significant differences in mean energy or nutrient intakes between the first and second FFQs, except for POI adjusted vitamin C intake ($p = 0.02$). The percentage difference between FFQ-1 and FFQ-2 ranged from 0% to 9.4% for unadjusted intakes, 0% to 10.5% for POI adjusted intakes, and 0% to 8.4% for EAT adjusted intakes.

Table 4.5: Cross-classification of children by quartiles of energy and nutrient intakes based on the FFQ and 5DDR

Nutrient	Unadjusted[a]		POI adjusted[a]		EAT adjusted[a]	
	% correctly classified[b]	% gross misclassification[c]	% correctly classified[b]	% gross misclassification[c]	% correctly classified[b]	% gross misclassification[c]
Energy (kJ)	39.9	4.6	38.6	4.6	38.6	3.3
Protein (g)	37.9	3.3	38.6	2.6	37.9	2.6
Total fat (g)	30.1	4.6	34.6	4.6	35.3	3.9
Carbohydrate (g)	48.4	3.9	41.6	1.9	43.8	1.3
Fibre (g)	42.7	1.2	42.5	1.3	43.1	0.7
Calcium (mg)	45.8	2.6	46.4	2.6	43.8	2.6
Iron (mg)	45.8	3.3	50.3	1.3	45.8	2.6
Zinc (mg)	34.6	4.6	34.6	3.9	37.9	3.9
Vitamin B_{12} (mg)	36.6	5.9	36.6	6.5	36.6	5.9
Vitamin C (mg)	39.2	5.9	37.3	2.6	46.4	5.2

[a] A two-part cross-check question was included at the beginning of both the fruit section and the vegetable section of the FFQ. 'Unadjusted' presents the energy and nutrient results before the cross-check questions are used to adjust energy and nutrient intakes in the FFQ. 'POI adjusted' results are for the FFQ adjusted only by the first part of the cross-check questions, and 'EAT adjusted' results are for the FFQ adjusted by both parts of the cross-check questions.

[b] % correctly classified = percentage of children with FFQ diets and 5DDR diets in the same quartile. If the two methods were completely unrelated, 25% correctly classified would be expected by chance.

[c] % gross misclassification = percentage of children with FFQ diets in the upper quartile and 5DDR diets in the lower quartile, and vice versa. If the two methods were completely unrelated 12.5% gross misclassification would be expected.

Table 4.6: Average daily intakes of energy and selected nutrients for the first and second FFQ administrations

Nutrient	Unadjusted[a]		POI adjusted[a]		EAT adjusted[a]	
	FFQ-1[b] Mean (95% CI)[d]	FFQ-2[c] Mean (95% CI)[d]	FFQ-1[b] Mean (95% CI)[d]	FFQ-2[c] Mean (95% CI)[d]	FFQ-1[b] Mean (95% CI)[d]	FFQ-2[c] Mean (95% CI)[d]
Energy (kJ)	5208 (4954,5476)	5204 (4958,5462)	4587 (4363,4823)	4632 (4418,4856)	5198 (4944,5465)	5282 (5022,5554)
Protein (g)	53 (50,56)	54 (52,58)	50 (48,53)	52 (49,55)	54 (51,57)	55 (52,58)
Total fat (g)	48 (45,51)	49 (46,52)	44 (42,47)	45 (43,48)	48 (45,51)	50 (47,53)
Carbohydrate (g)	149 (142,157)	146 (138,153)	124 (118,131)	122 (116,128)	148 (141,156)	148 (140,156)
Fibre (g)	13.3 (12.5,14.2)	13.1 (12.3,14.0)	9.5 (8.9,10.1)	9.5 (8.9,10.0)	13.4 (12.5,14.3)	13.4 (12.5,14.4)
Calcium (mg)	808 (746,876)	813 (755,876)	769 (706,838)	777 (720,840)	813 (751,879)	821 (764,883)
Iron (mg)	7.5 (7.0,8.0)	7.3 (6.8,7.9)	6.6 (6.1,7.0)	6.5 (6.0,6.90)	7.5 (7.0,8.0)	7.4 (6.9,8.0)
Zinc (mg)	7.2 (6.8,7.6)	7.2 (6.8,7.6)	6.7 (6.1,7.0)	6.7 (6.4,7.1)	7.2 (6.8,7.6)	7.2 (6.8,7.7)
Vitamin B_{12} (mg)	2.4 (2.3,2.6)	2.5 (2.3,2.6)	2.4 (2.2,2.6)	2.5 (2.3,2.6)	2.4 (2.3,2.6)	2.5 (2.3,2.6)
Vitamin C (mg)	77.4 (70.8, 84.6)	70.1 (64.1,76.7)	45.9 (41.7,50.5)	41.1 (37.2,45.4)	77.2 (70.0,85.1)	70.7 (63.9,78.3)

[a] A two-part cross-check question was included at the beginning of both the fruit section and the vegetable section of the FFQ. 'Unadjusted' presents the energy and nutrient results before the cross-check questions are used to adjust energy and nutrient intakes in the FFQ. 'POI adjusted' results are for the FFQ adjusted only by the first part of the cross-check questions, and 'EAT adjusted' results are for the FFQ adjusted by both parts of the cross-check questions.

[b] FFQ-1 was completed at baseline and [c] FFQ-2 was completed on average 5.0 weeks later (range: 3.6-8.7 weeks)

[d] Geometric mean and 95% confidence interval (CI)

4.3.2 Correlation Coefficients

Table 4.7 contains the intraclass correlation coefficients (ICC) used to assess the reproducibility between the first and second FFQ. The ICCs had the same mean (0.71) and range (0.65, 0.75) for unadjusted and POI adjusted data, but decreased slightly with EAT adjustment (mean 0.67; range 0.54, 0.72). Correlations were similar to those reported by D'Ambrosio (2012), with all unadjusted and POI adjusted correlations greater than 0.7, apart from vitamin B_{12} and vitamin C.

Table 4.7: Intraclass correlation coefficients for the repeated administrations of the FFQ, compared against the results reported by D'Ambrosio (2012)

Nutrient	D'Ambrosio (2012)[a]	Unadjusted[b]	POI adjusted[b]	EAT adjusted[b]
Energy (kJ)	0.63	0.72	0.72	0.68
Protein (g)	>0.7	0.71	0.70	0.71
Total fat (g)	>0.7	0.70	0.70	0.69
Carbohydrate (g)	>0.7	0.72	0.71	0.61
Fibre (g)	>0.7	0.71	0.73	0.64
Calcium (mg)	>0.7	0.75	0.75	0.74
Iron (mg)	>0.7	0.72	0.72	0.71
Zinc (mg)	-	0.71	0.70	0.72
Vitamin B_{12} (mg)	>0.7	0.69	0.69	0.69
Vitamin C (mg)	0.53	0.67	0.65	0.54

[a] Intraclass correlation coefficients between an FFQ and 24-hour recalls (12)

[b] A two-part cross-check question was included at the beginning of both the fruit section and the vegetable section of the FFQ. 'Unadjusted' presents the energy and nutrient results before the cross-check questions are used to adjust energy and nutrient intakes in the FFQ. 'POI adjusted' results are for the FFQ adjusted only by the first part of the cross-check questions, and 'EAT adjusted' results are for the FFQ adjusted by both parts of the cross-check questions.

5 Discussion

The EAT FFQ was designed to rank the nutrient intake of New Zealand toddlers aged 12 to 24 months based on their intake of selected nutrients over the past four weeks. The FFQ had adequate to good validity, with even better reproducibility, and performed similarly or better than previous toddler multi-nutrient FFQs. However, as is typically observed in the literature, our EAT FFQ overestimated energy and nutrient intakes compared to the diet record, and is therefore not an accurate measure of absolute intakes.

5.1 Validity

To assess the validity of the questionnaire to rank individuals according to their nutrient intake, correlation coefficients and cross-classification were used to compare the FFQ with the 5DDR. Although the primary aim of the questionnaire is to rank individuals, we were also interested in whether it is a suitable tool to assess absolute intakes. This was investigated using Bland-Altman statistics and mean dietary intakes. Unadjusted data were compared to the data adjusted by the fruit and vegetable cross-check questions, and then the results of this study compared with previous findings.

Previous research has related the overestimation of intake of fruits and vegetables to the number of questions asked on the FFQ (71). When a shorter list is used, less overestimation is found and vice versa. Questionnaires with longer lists can be adjusted for this overestimation by using a cross-check question, which asks for the total number of items eaten over a certain period of time (i.e. per week or per day). Each individual item in that section can then be adjusted using the total amount eaten. This has been shown to increase the agreement between the questionnaire and the reference method. It does, however, assume that each fruit and vegetable is over-reported to the same extent (59).

This study employed the use of cross-check questions in both the fruit and vegetable sections. The FFQ was originally designed with one cross-check question ‘How often has your toddler had vegetables in the past month’, however it was found that when participants answered this, they were counting the number of meals which contained fruits or vegetables, rather than the number of servings. A second part was then added to the cross-

check question to ensure participants counted every fruit and vegetable within a meal as a serving.

The correlation coefficients between the two methods generally improved with adjustment compared to the unadjusted data, with more correlations moving from the 'adequate' range (0.30-0.50) into the 'good' range (0.50-0.70) (7). Vitamin C had the largest increase in correlation, presumably because of its high concentration in fruit and vegetables. Vitamin B_{12} and calcium on the other hand, remained relatively unchanged due to their low concentrations in fruit and vegetables. Correlations were similar for EAT and POI adjusted data, with neither consistently higher than the other. For example, the correlation coefficient for zinc was 0.36 using POI adjusted data compared with 0.40 after EAT adjustment. In contrast, vitamin C was 0.57 (POI adjusted) compared to 0.51 (EAT adjusted). Other nutrients had closer correlations such as protein; 0.49 (POI adjusted) and 0.50 (EAT adjusted), presumably because fruit and vegetables are not major sources of protein and thus adjustment would not materially affect the outcomes.

Cross-classification was not affected by the cross-check questions, consistent with previous findings (59). The percentage correctly classified remained at approximately one-third to one-half of participants. An explanation for this could be that when the fruit and vegetable intakes were reduced, they were all reduced by similar amounts for each participant, so that although the mean intakes changed, the participants remained in similar quartiles.

Adjusting the data using the POI method generally decreased the nutrient means (i.e. carbohydrate: unadjusted 149g, POI adjusted 124g) causing the mean percentage agreement between the FFQ and 5DDR to be consistently closer to 100%. On the other hand, EAT adjustment had little effect on the unadjusted means, or the limits of agreement. The limits of agreement from the Bland-Altman analysis present a range within which 95% of the individual percentage differences (between the FFQ and 5DDR) fall and thus represent the agreement of the FFQ and 5DDR at the individual level (8). The unadjusted and EAT adjusted limits of agreement were wider than the POI adjusted data. For example, iron intake gave mean percentage agreement of 128%, with limits of agreement ranging from 58%-280% in the unadjusted data, which decreased to 112% (limits of agreement 106%,119%) for the POI adjusted data, whereas the EAT adjusted data were more comparable at 130% (59%,286%).

The validity of the adjusted data was higher than the unadjusted data according to the correlation coefficients and Bland-Altman statistics. There is little difference, however, between the POI and EAT adjusted intakes with respect to their ability to rank participants. However, POI adjusted data had a higher ability to assess the absolute intakes compared to the EAT adjusted data, and therefore it appears unnecessary to include the second part of the question when the FFQ is used in the future. POI adjusted data will be referred to for the remainder of this thesis.

The energy and nutrient correlation coefficients for the FFQ all lay within or above the ranges of the six previously reported toddler multi-nutrient FFQ validation studies (9–14). For example, the correlation between the FFQ and 5DDR for POI adjusted iron intake was 0.63, whereas in previous studies the range was 0.31-0.48. The correlations are also well within the 'adequate' and 'good' ranges of 0.3-0.70 as described by Willett (1998) (7).

This study validated the EAT FFQ, a modification of the FFQ used in the Southampton Women's Survey (SWS), validated by Marriott et al (2008). Compared to the validity of the original FFQ, the POI adjusted correlation coefficients were all higher than those of the original validation study, except for protein, zinc, and vitamin B_{12} (13).

The percentage correctly classified in the POI data ranged from 34.6% for total fat to 50.3% for iron, compared to an expected level of 25% by chance. The percentage of gross misclassification was low, all 6.5% or less, compared to an expected 12.5% by chance. This suggests that children would be ranked into similar quartiles using either the FFQ or 5DDR. The results were also similar to, if not higher than, those reported in the two studies by Andersen et al. (2003, 2004), who also used quartiles, and correctly classified 22-52% (9) and 29-37% (10) of participants, respectively, when looking at energy and the nine nutrients assessed in this study. These results suggest that our EAT questionnaire is capable of identifying children with extremes of dietary intake.

Mean POI adjusted energy and nutrient intakes were overestimated by this FFQ compared to the 5DDR (33% for energy; 12-60% for nutrients), except for vitamin C intake which had good agreement with the 5DDR. This is common among FFQs, with energy overestimated in five out of the other six toddler multi-nutrient FFQ validation studies, by 10-70% ((9,10,12–14)).

Two of the biggest challenges for an FFQ for toddlers that may cause overestimation are the inaccuracies surrounding portion size estimation (8), and leftovers. Toddlers eat small portions, and tend to leave food uneaten, which is not necessarily left on the plate. This study took two unique approaches to these issues. Firstly, information was collected on both the amount offered, and then the amount actually eaten, in both the FFQ and 5DDR. This ensured the participant remembered to take into consideration the uneaten portion. The second approach used in the FFQ was to ask participants to describe the portion sizes of foods in units of the child's 'palm volume'. A child's palm is very small and therefore possibly closer in size to the portions of food eaten than other measurements commonly used, such as a cups and bowls. Another advantage is that the child's palm is always seen next to their food, and so may be easier for the caregiver to visualise than the cups or bowls which may either not be used, or be used in a different size to those used in research studies. These techniques were used to help decrease the overestimation of the FFQ. However, in practice the FFQ still overestimated energy and all nutrient intakes by a significant amount, except for POI adjusted vitamin C intake. It was beyond the scope of this thesis to assess the appropriateness of using palm volume to quantify intake. However it would appear from the results of this validation study that it is an acceptable portion size estimate as the validity analysis showed the FFQ was an appropriate method for ranking toddlers by their nutrient intake.

The limits of agreement from the Bland-Altman analysis showed that while the majority of participants overestimated nutrient intakes, some participants underestimated them. For example, while the mean percentage agreement for unadjusted carbohydrate is 134 (a 34% overestimate), the limits of agreement were 79%-227%. This shows the wide variation within the study, with some participants underestimating their child's intake compared to the 5DDR by up to 21% while others overestimated the intake by 127%.

There are no set recommendations for what can be categorised as adequate results from Bland-Altman analysis, and therefore the only way to judge the results is by comparison with previous studies. The Bland-Altman statistics, however, have only been used in three toddler multi-nutrient validation studies (9,12,13), and only one, the validation study for the FFQ on which this one is based, reported the mean percentage agreement and the limits of agreement (13).

The Bland-Altman statistics were comparable or higher for a number of nutrients in this study, with the SWS FFQ mean percentage agreement ranging from 112% (calcium)-146% (iron), compared to POI adjusted values of 112% (carbohydrate)-160% (zinc), when vitamin C was excluded from both studies (13). Vitamin C was overestimated by 60% in the SWS FFQ compared to the agreement found in this study, with POI adjusted vitamin C having a mean percentage agreement of 93% and the 95% CI including 100% (95% CI 85%-103%). The improvement in our study relative to the SWS FFQ for estimates of vitamin C can be explained by our adjustment of the fruit and vegetable sections using the cross-check questions, a modification of the original questionnaire. Fruits and vegetables have a high vitamin C content, which explains why adjustment of their intake has a greater effect on vitamin C estimates than it does on estimates for the other nutrients.

5.2 Reproducibility

Intraclass correlation coefficients (ICC) were used to assess the reproducibility of the questionnaire. The repeatability of the FFQ was consistently high irrespective of the level of adjustment with the cross-check questions. Although the EAT adjusted results tended to produce lower ICCs between the two FFQ administrations than the unadjusted and POI adjusted data, they were still relatively high, and all were greater than 0.61 with one exception (vitamin C). The POI adjusted ICCs ranged from 0.65-0.75, all in the 'reasonable/good' to 'very good' range (39). This is also consistent with a review article in all age groups, which found reproducibility correlations were usually between 0.5 and 0.7 for most nutrients (25).

There is only one other toddler multi-nutrient FFQ study that has assessed reproducibility of nutrient intakes, and they did not report correlations for all nutrients (12). Nevertheless, using the information provided, the current study achieved very similar results with only vitamin B_{12} (0.69 unadjusted and POI adjusted) and vitamin C (0.67 unadjusted, 0.65 POI adjusted) intakes having a correlation coefficient less than 0.70, compared to only energy and vitamin C with correlation coefficients less than 0.70 in the study by D'Ambrosio et al (12).

The high reproducibility in the current study could be in part explained by the short time frame between the first and second questionnaire. The risk of administering the two FFQs only four weeks apart (although we ended up with a mean of five weeks) is that participants

may be able to remember their answers from the first FFQ (7). Adult studies have found correlations tend to be higher when the repeat FFQs are administered one month or less apart, due to a 'learning' effect (25). The period the FFQ assesses is also important as our questionnaire was only assessing the past four weeks of habitual intake, compared to the usual one year in adult studies (25). It was decided for this study that the time frame of four weeks was required because young children's dietary patterns change quickly, with caregivers often commenting at the second visit that their child had changed their food preferences since the first visit. This appears not to have impacted negatively on the FFQ reproducibility, possibly because the short time frame only allowed for small changes in diet. A longer period between questionnaires may have resulted in a true change of diet, reducing the agreement between questionnaires. It is also possible that a child may change their preferences for foods with similar nutrient content. For example if a child usually liked to eat carrots with their dinner, and then decided they preferred pumpkin, the similar nutrient profile would not alter the reproducibility of the results. Seasonal affects would also be very minimal or non-existent in this study due to the short time frame covered.

5.3 Study Strengths and Limitations

Dietary assessment, particularly in toddlers, is a difficult task with many complications to consider. Issues including generalisability of the study results, reporting by secondary carers, rate of sickness in toddlers, accuracy of the reference method and nutrient lines, and design of the study are discussed in the following section, in relation to their effect on the results.

A challenge facing all FFQ validation studies is recruiting participants who are representative of the population the FFQ will be used in. Our study used a convenience sample of participants recruited mainly through newspaper advertisements and posters. It is likely that the participants who volunteered for our study were more motivated than the general public, demonstrated through our high completion rate of 96%. The participants were also very interested in finding out about their child's nutrient intake. This could indicate that the reason the caregivers volunteered for the study was driven by their interest in nutrition, potentially biasing our sample group.

The anthropometric data showed the children in the study were heavier and taller for their age compared to other toddlers in the reference populations (68). However, this isn't surprising as previous research has shown New Zealand toddlers to have BMI-for-age z

scores to be higher than the growth standard (72). There was some ethnic diversity in our study, however not at levels representative of the New Zealand population. Data collected in the 2006 census found 25.5% of children 0-4 years were identified as Māori and 13.7% as Pacific Island, compared to this study with 7.8% and 2.6% for Māori and Pacific Island respectively (73). This study also over-represented participants from the lower and middle deciles of deprivation (42% in deciles 1-3, 46.9% in deciles 4-7; when 30% and 40% would be expected respectively), and under-represented in the higher deciles of deprivation (11.2% in deciles 8-10; when 30% would be expected). Further investigation would be required if the FFQ was to be used in minority populations such as communities with high deprivation, or those containing a high proportion of Māori or Pacific Island people.

Only 16% of children in the study were fed solely by the primary caregiver completing the questionnaire. This means that 84% of the primary caregivers were not present at every meal to see what their child was being fed, and to see how much they ate or left over. A total of 38% of these caregivers did not know about all the food that their child was fed (mean 6.3 meals/week, range 1-25 meals/week).

This will decrease the accuracy with which the caregivers were able to report on their child's dietary intake in the FFQ. It is important to note here that, although the primary caregivers were not with their child at all times, it appears that FFQs are less affected by this than other dietary assessment methods such as the 24-hour recall. Previous research has found parents who spent greater than 4.5 hours per day away from their child had difficulties in reporting any of their child's dietary intake for that section of the day, when using a 24-hour recall (22). This is compared to a study using FFQs where no difference was found in the agreement between the FFQ and reference method for caregivers who were with their child at all feeding times, versus the diets of children who spent time with other carers (14). Therefore, not being present at all meals may not have affected the participant's abilities to complete the FFQ in this study. This would mean any discrepancy between the FFQ and diet record cannot be explained by this, as secondary carers were also asked to complete the diet record when the child was in their care. If the secondary carers did not complete the diet record to the standard required (i.e. did not record all foods eaten), this could explain the apparent overestimation of the FFQ which would actually be due to an underestimation of dietary intake by the diet records.

Participants were provided with spare pages to give secondary carers during the diet record so that foods eaten during their care could be recorded. These secondary carers, however, rarely completed the record with the accuracy required. While some carers filled out the diary with the appropriate details, others tended to be less specific about the foods and amounts eaten. In particular, when the child went to a day care centre the supervisors tended to write down the name of the meal (e.g. Spaghetti Bolognese), possibly with a few key ingredients (e.g. mince, tomato, pasta, cheese), and then weigh the entire meal as one. Although not as common, this was also done by the primary caregivers at times. Asking for each ingredient does increase the participant burden; however it reduces the errors involved when a researcher has to estimate the amounts of each individual food in the meal.

It is not possible to have a dietary assessment method that measures usual dietary intake with complete accuracy. The weighed diet record is the most appropriate method to use with toddlers when validating an FFQ as it allows for portions to be directly measured and it has errors which are unrelated to those in an FFQ (15). A strength of this study was the ability to collect five non-consecutive days of weighed diet record data over four weeks. This provided a good estimate of the toddlers' usual dietary intake, with all days of the week represented across the sample (31% of records were weekend days). Only three of the six previous toddler multi-nutrient FFQ validation studies used a weighed diet record as the reference method (9,10,13), the rest using 24-hour recalls (11,12,14). As these rely on memory to recall the foods eaten and estimates of the portion sizes, as does an FFQ, they are not as appropriate as the weighed diet record (17).

Nevertheless, the diet record still contains errors which could attenuate the agreement found between the FFQ and 5DDR. A limitation of the diet record is the burden it places on the participant in weighing the foods, which can cause caregivers either to misreport, or to choose foods that are easier to weigh. For example, a caregiver mentioned in the second interview that they normally would not give their child tinned infant food, however, they had especially gone out and bought tinned food to feed their child on the diet record days. Although not many participants may have gone to quite the extent of specially buying easy to weigh foods, it is quite conceivable that they may have avoided feeding meals that were difficult to weigh on diet record days, even though they had been asked not to change their child's diet. These changes in diet will mean the diet record is less representative of their habitual diet, but as the recording days were only once a week, are unlikely to have had a

significant impact on their responses to the FFQ, therefore reducing the agreement between the two methods. This effect will, however, depend on the foods that are chosen to replace the 'difficult to weigh' foods. It is likely that the caregivers would still have chosen foods that are similar to what they would normally feed their child. If they had a similar nutrient content, the substitutions would not affect the mean daily nutrient intakes significantly.

In the current study, a child was sick on 18% of all the recorded diet record days, and on 64% of these days the parent considered that the child's dietary intake was affected (93% of the these affected days, the parent considered that there was a decrease in intake). It was decided not to exclude these days as although their dietary intake was altered, it is usual for toddlers to have frequent illnesses, and thus should be seen as 'normal' variation. A study conducted in 1-2 year old New Zealand children in 2004 found that toddlers had 3.1 sick days on average over a four week period (74). Our study found the children were sick on 0.9 days out of five, over the four weeks of diet recording, however the data are difficult to compare, as we did not ask about the days on which the diet record was not being kept. Irrespective of how comparable the data are, both studies show that toddlers are often sick, and this will regularly affect their dietary intake. Therefore, the foods eaten on the 'sick days' are a part of their usual eating patterns. How this affects the results of the validity analysis depends on whether the parent remembered to incorporate the days of affected dietary intake into their answers for the FFQ. It is likely this will have attenuated the validity between the FFQ and 5DDR if they counted these days as anomalies and did not consider them in their FFQ answers and this may account for some of the overestimation of intakes.

Due to the limitations of dietary assessment methods, validation studies often use biomarkers as an additional indicator of habitual intake alongside diet records (17). They are less prone to errors and under- and over-reporting as they are independent of reported dietary intake. Biomarkers, however, are invasive and therefore less appropriate in a toddler population compared to a study in adults. They are also nutrient-specific as only energy and some nutrients including protein and vitamin C can be assessed with this technique (8). As with any dietary assessment method, they also contain errors, as it is difficult to ascertain the true relationship between the biochemical marker and usual dietary intake. Factors such as digestion, absorption, metabolism and excretion all complicate the relationship (17). It is also recommended that multiple measures are taken, to take into account diurnal variation

in concentration, further increasing the participant burden (17). It was for these reasons that biomarkers were not included in this study.

FOODfiles 2010 is the most recent nutrient database available containing New Zealand foods (60). New food products are always coming out on the market however, and FOODfiles did not contain some of the foods eaten by the participants. This required substitutions to be made decreasing the accuracy of the diet records, but also potentially increasing agreement with the questionnaire. When a particular food was not present in the database (e.g. 'Tasti Milkies muffin bars'), a composite food item was often chosen (e.g. 'Cereal soft bar, assorted flavours'). Composite foods are food items which use an average nutrient line made by combining multiple foods. The composite food items were also often used in the FFQ nutrient lines, and so in using these averaged nutrient lines in both methods, the agreement between the two methods is likely to have been strengthened.

FOODfiles 2010 did not contain any information on the nutrient content of baby and infant foods. As a result, the baby and infant foods reported in the diet records were analysed using the nutrient lines from the 2006 version of the database. There have been some concerns about the reliability of these data and they may be outdated. Unfortunately it was outside the scope of this thesis to modify foods in the diet record entry programme 'Kai-culator' to better reflect the composition of these infant and toddler foods. The FFQ, however, used recipes developed by the candidate to match the nutrient content and recipe list on the back of the actual products. This will affect a number of participants, lowering the relative validity of the questionnaire, as 46% of participants reported consuming on average 1 serving per day of baby or infant food in their diet record.

The majority of nutrient lines in the FFQ were weighted using data from a previous New Zealand study in toddlers (58). This data was the best information available at the time, and although the data were 14 years old, it should still provide a better estimate than arbitrary decisions. The way the FFQ questions were grouped should also lower the effect of any changes in food habits since the data were collected. For example, question 35 on the FFQ asked about 'other meat' (which was any meat other than chicken, fish and most processed meats which had already been asked earlier). Although the overall consumption of meat may have increased (75), it is unlikely that the proportions of lamb, pork and beef eaten have changed significantly. Detailed information was not available, however, on the

different types of dried fruit eaten for example. The study also did not provide information on cooking methods such as how often baked, mashed or boiled potatoes were eaten. As there was no formal information available to guide these decisions, subjective decisions based on 'common sense' had to be made. It may be the case that there are weightings within the nutrient lines that are not the best fit for this population, decreasing the validity of the questionnaire.

Breast milk was excluded from both the FFQ and 5DDR results. As 29% of children in the study were breastfed, this is a large proportion of participants who would have had a significant part of their daily nutrient intake missing, meaning the child's entire daily intake was not being assessed. However, to gain accurate data on the amount of breastmilk consumed, and its nutrient content, would have been very difficult and required test weighing, and breast milk sample collection and analysis. This was not feasible in this study. As breast milk was excluded from both methods, the agreement between the two methods was not altered. Supplement data were also excluded from the analysis as they were not collected in the FFQ.

A strength of this study was its design, which allowed both validity and reproducibility to be assessed in parallel. Only one other toddler multi-nutrient FFQ has been assessed for reproducibility, a quality that should be assessed as high validity does not necessarily infer good reproducibility. The second FFQ was completed after the diet record, so the participants' answers could have been influenced by their heightened awareness of foods eaten, due to recording foods over the previous month. This is why a random sample of first and second FFQs was used to determine validity. However, the high agreement between the two FFQs suggests the diet record did not influence the second FFQ (7).

Another strength of this study was its ability to recruit a large sample size. It is advised that validity studies have a sample size between 100 and 200 (17), yet only two previous toddler multi-nutrient FFQ validation studies had at least 100 participants (10,11). Our study was able to recruit 153 in the validity analysis and 152 in the reproducibility analysis. This increases the likelihood of our study being a true representation of the population we studied.

5.4 Future Research

Adjustment of the nutrient intake data for energy intake would be beneficial. Energy adjustment allows the intake of a nutrient to be related, independent of energy intake, with a particular outcome (17), so it would be useful to see how well the FFQ performs following energy adjustment. It is also often found that the correlation between an FFQ and its reference method increases following energy adjustment, particularly for energy-yielding nutrients (76).

It would also be beneficial to conduct further research into using the child's 'palm volume' as a way of measuring portion size. In particular, whether palm size is easier for participants to use than other available methods. Portion size estimation is an area in which more research is required overall, as all the currently available methods are prone to errors (77).

Following on from research into the use of palm volume as a method of determining portion size, it would be useful to determine whether it is possible to use an average palm size for participants based on their age. This would allow the questionnaire to be used in non-interviewer situations, expanding the potential geographical reach of the questionnaire. Using a mailed version of the questionnaire would reduce costs significantly and a web-based version would have even more benefits (7). A web-based version would not only reduce the printing costs, but also the costs involved with entering the data, and errors related to coding. These changes however, would require the FFQ to be re-validated to ensure its appropriateness.

6 Conclusion

There are currently only six published food frequency questionnaires (FFQs) assessed for their relative validity when determining multi-nutrient intake in toddlers between 12 and 24 months of age, none of which have been validated for use in New Zealand (9–14).

The EAT FFQ is a 98-item, multi-nutrient FFQ modified from a questionnaire used in the Southampton Women's Survey. It was designed to rank New Zealand toddlers aged 12-24 months based on their average daily energy and nutrient intake.

This study has shown that the EAT FFQ has adequate to good relative validity when compared with a five-day weighed diet record, and has high reproducibility when measured over one month. Adjusting the fruit and vegetable sections for total intake, using the POI technique, improved the questionnaire's relative validity further. The questionnaire is not, however, suitable to assess absolute intakes, except for POI adjusted vitamin C. This is consistent with most other FFQs, which characteristically do not perform well at the individual level (8). The questionnaire is therefore acceptable for ranking the diets of New Zealand toddlers, and for identifying toddlers at either end of the nutrient intake distribution.

7 References

1. Lytle LA. Nutritional issues for adolescents. Journal of the American Dietetic Association 2002;102(3):S8–12.

2. Ministry of Health. A Portrait of Health. Key Results of the 2006/07 New Zealand Health Survey. Wellington; 2008.

3. Organization WH. Obesity: Preventing and Managing the Global Epidemic. Report on a WHO Consultation. Geneva, Switzerland; 2004.

4. Soh P, Ferguson E, McKenzie J, Homs M, Gibson R. Iron deficiency and risk factors for lower iron stores in 6-24-month-old New Zealanders. European Journal of Clinical Nutrition 2004;58(1):71–9.

5. Morgan E, Heath A-LM, Szymlek-Gay EA, Gibson RS, Gray AR, Bailey KB, et al. Red meat and a fortified manufactured toddler milk drink increase dietary zinc intakes without affecting zinc satus of New Zealand toddlers. The Journal of Nutritional Epidemiology 2010;140:2221–6.

6. Emmett P. Workshop 2: The use of surrogate reporters in the assessment of dietary intake. European Journal of Clinical Nutrition 2009;63:S78–S79.

7. Willett W. Nutritional Epidemiology. Second Ed. New York: Oxford University Press; 1998.

8. Gibson RS. Principles of Nutritional Assessment. Second Ed. New York: Oxford University Press; 2005.

9. Andersen L, Lande B, Arsky G, Trygg K. Validation of a semi-quantitative food-frequency questionnaire used among 12-month-old Norwegian infants. European Journal of Clinical Nutrition 2003;57(8):881–8.

10. Andersen L, Lande B, Trygg K, Hay G. Validation of a semi-quantitative food-frequency questionnaire used among 2-year-old Norwegian children. Public Health Nutrition 2004;7(06):757–64.

11. Blum RE, Wei EK, Rockett H, Langeliers JD, Leppert J, Gardner JD, et al. Validation of a food frequency questionnaire in Native American and Caucasian children 1 to 5 years of age. Matern Child Health Journal 1999;3(3):167–72.

12. D'Ambrosio A, Tiessen A, Simpson JR. Development of a food frequency questionnaire for toddlers of Low-German-Speaking Mennonites from Mexico. Canadian Journal of Dietetic Practice and Research 2012;73(1):40–4.

13. Marriott LD, Inskip HM, Borland SE, Godfrey KM, Law CM, Robinson SM, et al. What do babies eat? Evaluation of a food frequency questionnaire to assess the diets of infants aged 12 months. Public Health Nutriton 2008;12(7):967–72.

14. Parrish L, Marshall J, Krebs N, Rewers M, Norris J. Validation of a food frequency questionnaire in preschool children. Epidemiology 2003;14(2):213–7.

15. Margetts BM, Nelson M. Design Concepts in Nutritional Epidemiology. Second Ed. New York: Oxford Medical Publications; 2010.

16. Cameron M, Van Straveren W. Manual on Methodology for Food Consumption Studies. Oxford: Oxford University Press; 1988.

17. Cade J, Thompson R, Burley V, Warm D. Development, validation and utilisation of food-frequency questionnaires - a review. Public Health Nutr. 2002;5(4):567–87.

18. Kristal AR, Peters U, Potter JD. Is it time to abandon the food frequency questionnaire? Cancer Epidemiology, Biomarkers & Prevention 2005;14(12):2826–8.

19. Bowens J, Hertzler A, Hull S. Preschoolers' reporting of food habits. Journal of the American Dietetic Association 1993;93(10):1159–61.

20. Klesges R, Klesges L, Brown G, Frank G. Validation of the 24-hour recall in preschool children. Journal of the American Dietetic Association 1987;87(10):1383–5.

21. Eck LH, Klesges R, Hanson C. Recall of a child's intake from one meal: are parents accurate? Journal of the American Dietetic Association 1989;89(6):784–9.

22. Baranowski T, Sprague D, Baranowski J, Harrison J. Accuracy of maternal dietary recall for preschool children. Journal of the American Dietetic Association 1991;91(6):669–74.

23. Bradburn N, Rips L, Shevell S. Answering autobiographical questions: The impact of memory and inference on surveys. Science 1987;236(4798):157–61.

24. Dubois S. Accuracy of visual estimates of plate waste in the determination of food consumption. Journal of the American Dietetic Association 1990;90:382–7.

25. Cade J, Burley V, Warm D, Thompson R, Margetts B. Food-frequency questionnaires: a review of their design, validation and utilisation. Nutrition Research Reviews 2004;17(1):5–22.

26. Molag ML, De Vries JHM, Ocké MC, Dagnelie PC, Van den Brandt P a, Jansen MCJF, et al. Design characteristics of food frequency questionnaires in relation to their validity. American Journal of Epidemiology 2007;166(12):1468–78.

27. Subar AF, Thompson F, Smith A, Jobe J, Ziegler R, Potischman N, et al. Improving food frequency questionnaires: A qualitative approach using cognitive interviewing. Journal of the American Dietetic Association 1995;95:781–8.

28. Nelson M, Atkinson M, Darbyshire S. Food photography II: use of food photographs for estimating portion size and the nutrient content of meals. British Journal of Nutrition 1996;76:31–49.

29. Young LR, Nestle MS. Portion sizes in dietary assessment: issues and policy implications. Nutrition Reviews 1995;53(6):149–58.

30. Nelson M, Atkinson M, Darbyshire S. Food Photography I: the perception of food portion size from photographs. British Journal of Nutrition 1994;72(5):649–63.

31. Faggiano F, Vineis P, Cravanzola D, Pisani P, Xompero G, Riboli E, et al. Validation of a method for the estimation of food portion size. Epidemiology 1992;3:379–82.

32. Willett W. Future directions in the development of food-frequency questionnaires. American Journal of Clinical Nutrition 1994;59:171S–9S.

33. Samet JM, Humble CG, Skipper BE. Alternatives in the collection and analysis of food frequency interview data. American Journal of Epidemiology 1984;120(4):572–81.

34. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Physical Therapy 2005;85:257–68.

35. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;i:307–10.

36. Thompson FE, Byers T. Dietary assessment resource manual. The Journal of Nutrition 1994;124(11):2245S–317S.

37. Livingstone M, Robson P, Wallace J. Issues in dietary intake assessment of children and adolescents. The British Journal of Nutrition 2004;92(S2):S213–22.

38. Henríquez-Sánchez P, Sánchez-Villegas A, Doreste-Alonso J, Ortiz-Andrellucchi A, Pfrimer K, Serra-Majem L. Dietary assessment methods for micronutrient intake: a systematic review on vitamins. The British Journal of Nutrition. 2009;102:S10–37.

39. Serra-Majem L, Frost Andersen L, Henríque-Sánchez P, Doreste-Alonso J, Sánchez-Villegas A, Ortiz-Andrelluchi A, et al. Evaluating the quality of dietary intake validation studies. The British Journal of Nutrition 2009;102:S3–9.

40. Serra-Majem L, Pfrimer K, Doreste-Alonso J, Ribas-Barba L, Sánchez-Villegas A, Ortiz-Andrellucchi A, et al. Dietary assessment methods for intakes of iron, calcium, selenium, zinc and iodine. The British Journal of Nutrition 2009;102:S38–55.

41. Lanigan J, Wells J, Lawson M, Cole T, Lucas A. Number of days needed to assess energy and nutrient intake in infants and young children between 6 months and 2 years of age. European Journal of Clinical Nutrition 2004;58(5):745–50.

42. Stram DO, Longnecker MP, Shames L, Kolonel LN, Wilkens LR, Pike MC, et al. Cost-efficient design of a diet validation study. American Journal of Epidemiology 1995;142(3):353–62.

43. Hartman A, Brown C, Palmgren J, Pietinen P, Verkasalo M, Myer D, et al. Variability in nutrient and food intakes among older middle-aged men. American Journal of Epidemiology 1990;132(5):999–1012.

44. Livingstone MBE, Robson PJ. Measurement of dietary intake in children. Proceedings of the Nutrition Society 2007;59(02):279–93.

45. Masson LF, McNeill G, Tomany JO, Simpson JA, Peace HS, Wei L, et al. Statistical approaches for assessing the relative validity of a food-frequency questionnaire: use of correlation coefficients and the kappa statistic. Public Health Nutrition 2003;6(3):313–21.

46. Lee J. Alternate approaches for quantifying aggregate and individual agreements between two methods for assessing dietary intakes. The American Journal of Clinical Nutrition 1980;33:956–64.

47. Block G, Hartman AM. Issues in reproducibility and validity of dietary studies. The American Journal of Clinical Nutrition 1989;50:1133–8.

48. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. The Statistician 1983;32:307–17.

49. Altman D, Bland J. Commentary on quantifying agreement between two methods of measurement. Clinical Chemistry 2002;48(5):801–2.

50. MacIntyre U, Venter C, Vorster H, Steyn H. A combination of statistical methods for the analysis of the relative validation data of the quantitative food frequency questionnaire used in the THUSA study. Public Health Nutrition 2007;4(01):45–51.

51. Bland JM, Altman DG. Measuring agreement in method comparison studies. Statistical Methods in Medical Research 1999;8(2):135–60.

52. Cohen J. Weighted Kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. Psychological Bulletin 1968;70(4):213–20.

53. Weir J. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. Journal of Strength and Conditioning Research 2005;19(1):231–40.

54. Atkinson G, Nevill A. Statistical methods for assessing measuring error (reliability) in variables relevant to sports medicine. Sports Medicine 1998;26(4):217–38.

55. Hopkins WG. Measures of reliability in sports medicine and science. Sports Medicine 2000;30(1):1–15.

56. Klohe DM, Clarke KK, George GC, Milani TJ, Hanss-Nuss H, Freeland-Graves J. Relative validity and reliability of a food frequency questionnaire for a triethnic population of 1-year-old to 3-year-old children from low-income families. Journal of the American Dietetic Association 2005;105(5):727–34.

57. Hartman A, Brown C, Palmgren J, Pietinen P, Verkasalo M, Myer D, et al. Variability in nutrient and food intakes among older middle-aged men. American Journal of Epidemiology 1990;132(6):999–1012.

58. Szymlek-Gay E, Ferguson E, Heath A-L, Fleming E. Quantities of foods consumed by 12- to 24-month-old New Zealand children. Nutrition & Dietetics 2010;67:244–50.

59. Calvert C, Cade J, Barrett J, Woodhouse A, UKWCS Steering Group. Using cross-check questions to address the problem of mis-reporting of specific food groups on food frequency questionnaires. European Journal of Clinical Nutrition 1997;51:708–12.

60. New Zealand Istitute for Crop & Food Research. FOODfiles: Data files of the New Zealand Food Composition Database. Palmerston North: New Zealand Institute for Crop & Food Research; 2010.

61. New Zealand Istitute for Crop & Food Research. FOODfiles: Data files of the New Zealand Food Composition Database. Palmerston North: New Zealand Institute for Crop & Food Research; 2006.

62. Ministry of Health. Eating for healthy babies and toddlers from birth to 2 years old [Internet]. Wellington, New Zealand; 2008. Available from: www.healthed.govt.nz

63. Ministry of Health. Eating for healthy children aged 2 to 12 [Internet]. Wellington, New Zealand; 2010. Available from: www.healthed.govt.nz

64. Statistics New Zealand. New Zealand census of population and dwelling [Internet]. 2006. Available from: http://www.stats.govt.nz/Census/about-2006-census/2006-census-questionnaires.aspx

65. Onis M, Onvango A, Van den Broeck J, Chumlea W, Martorell R. Measurement and standardization protocols for anthropometry used in the construction of a new international growth reference. Food Nutrition Bulletin 2004;25:S27–36.

66. Food Standards Australia New Zealand. Nutrition information requirements [Internet]. 2002 p. 97–110. Available from: http://www.foodstandards.gov.au/foodstandards/userguides/nutritioninformation1406.cfm

67. Taylor B, Heath A-L, Galland B, Gray A, Lawrence J, Sayers R, et al. Prevention of Overweight in Infancy (POI.nz) study: a randomised controlled trial of sleep, food and activity interventions for preventing overweight from birth. BMC Public Health 2011;11:942.

68. Ministry of Health. Growth charts [Internet]. Wellington; 2010. Available from: http://www.health.govt.nz/our-work/life-stages/child-health/well-child-services/growth-charts

69. Ministry of Health. Fact sheet 2: About the NZ-WHO groth charts [Internet]. New Zealand; 2010. Available from: http://www.health.govt.nz/our-work/life-stages/child-health/well-child-services/growth-charts

70. Salmond C, Crampton P, Atkinson J. NZDep2006 Index of Deprivation User's Manual. Wellington; 2007.

71. Krebs-Smith S, Heimendinger J, Subar A, Patterson B, Pivonka E. Using food frequency questionnaires to estimate fruit and vegetable intake: Association between the number of questions and total intakes. Journal of Nutrition Education 1995;27(2):80–5.

72. Szymlek-gay EA, Ferguson EL, Heath AM, Gray AR, Gibson RS. Food-based strategies improve iron status in toddlers : a randomized controlled trial. American Journal of Clinical Nutrition 2009;1541–51.

73. Statistics New Zealand. Estimated national ethnic population by age and sex at 30 June 1996, 2001 & 2006 [Internet]. 2006. Available from: http://wdmzpub01.stats.govt.nz/wds/TableViewer/tableView.aspx

74. Cross A, Heath A-L, Ferguson E, Gray A, Szymlek-Gay E. Rates of common communicable illnesses in non-anaemic 12-24 month old South Island, New Zealand children. The New Zealand Medical Journal 2009;122(1290):24–35.

75. Chang J. Meat production and consumption continue to grow [Internet]. Worldwatch Institute. Available from: http://vitalsigns.worldwatch.org/vs-trend/meat-production-and-consumption-continue-grow-0

76. Bingham S. Using biochemical markers to assess the validity of prospective dietary assessment methods and the affect of energy adjustment. The American Journal of Clinical Nutrition 1997;65(4):1130S.

77. Cypel Y, Guenther P, Petot G. Validity of portion-size measurement aids: A review. Journal of American Dietetics Association. Journal of American Dietetics Association 1997;97:289–92.

8 Appendices

A EAT Food Frequency Questionnaire

B Ethics Approval

C Māori Consultation

D Advertising Protocol

E Study Poster

F Nutrient Analysis Outline

G Recruitment Protocol

H Information Letter, Information Sheet, and Consent Form

I First Visit Protocol

J Second Visit Protocol

K Demographic Questionnaire

L Anthropometric Protocol

M Diet Record

N Instructions for Scales

O Reminder Protocol

P Take-a-way Food Sheet

Q Secondary Carers Form

R Food Frequency Questionnaire Entry Protocol