Review Article

Evaluation of tools used to measure calcium and/or dairy consumption in children and adolescents

Anthea Magarey*, Alison Yaxley, Kylie Markow, Lauren Baulderstone and Michelle Miller
Nutrition and Dietetics, Flinders University, GPO Box 2100, Adelaide, SA 5001, Australia

Abstract

Objective: To identify and critique tools that assess Ca and/or dairy intake in children to ascertain the most accurate and reliable tools available.
Design: A systematic review of the literature was conducted using defined inclusion and exclusion criteria. Articles were included on the basis that they reported on a tool measuring Ca and/or dairy intake in children in Western countries and reported on originally developed tools or tested the validity or reliability of existing tools. Defined criteria for reporting reliability and validity properties were applied.
Setting: Studies in Western countries.
Subjects: Children.
Results: Eighteen papers reporting on two tools that assessed dairy intake, ten that assessed Ca intake and five that assessed both dairy and Ca were identified. An examination of tool testing revealed high reliance on lower-order tests such as correlation and failure to differentiate between statistical and clinically meaningful significance. Only half of the tools were tested for reliability and results indicated that only one Ca tool and one dairy tool were reliable. Validation studies showed acceptable levels of agreement (<100 mg difference) and/or sensitivity (62–83%) and specificity (55–77%) in three Ca tools. With reference to the testing methodology and results, no tools were considered both valid and reliable for the assessment of dairy intake and only one tool proved valid and reliable for the assessment of Ca intake.
Conclusions: These results clearly indicate the need for development and rigorous testing of tools to assess Ca and/or dairy intake in children and adolescents.

Bone health benefits derived from consuming dairy products, the largest contributors of Ca to the Western diet[1], are well documented in the scientific literature[2]. There is some evidence for the role of dairy food consumption in reducing the risk of chronic disease, i.e. obesity and hypertension, but further studies are required to confirm a causal relationship[3]. An adequate Ca intake is particularly important during childhood and adolescence to optimise peak bone mass and consequently is a readily modifiable factor in reducing osteoporosis risk[4]. However, intake studies suggest that while younger children appear to consume appropriate amounts of dairy foods and are meeting the recommendations for Ca intake, older children and adolescents tend to consume less dairy foods than their younger counterparts and thus fail to meet recommended intakes of Ca[5–7].

In order to address shortfalls in Ca intake in a population, accurate assessment of Ca intake is required to identify those with suboptimal intakes, thus enabling public health strategies to be targeted appropriately. As dairy foods are the principal source of Ca in the Western diet, assessment of dairy intake can potentially provide a reasonable estimate of Ca intake. For population-wide assessment of Ca intake, or dairy as a proxy for Ca, a dietary intake tool that is short, reliable, relatively simple and inexpensive to administer is ideal.

For a tool to be recommended for use it needs to be reliable and valid. Reliable tools produce consistent results when performed under similar circumstances, e.g. at repeat administrations and/or when conducted by different researchers[8]. Criterion validity identifies the ability of a tool to accurately measure what it proposes to

Corresponding author: Email anthea.magarey@flinders.edu.au

measure, usually determined by how closely the results match those of a reference test[8,9]. The present paper is the first of two reviews with the overall aims to identify published tools that estimate Ca and/or dairy intake in children, adolescents and adults and to assess the testing of tool properties in order to recommend a tool(s) for use. The current paper focuses on tools developed for use with children and adolescents.

Methods

A comprehensive search was completed to identify existing tools to measure dairy and/or Ca intake. The search was conducted using the databases MEDLINE, Scopus, Ovid, Informit and Web of Knowledge, with the keywords 'calcium', 'dairy', 'milk', 'diet', 'nutrition' and 'food', combined with 'tool', 'questionnaire', 'FFQ', 'measurement', 'assessment', 'evaluation' and 'analysis'. The search was not limited by dates, but databases were searched from their year of inception, the earliest being 1948 in the case of MEDLINE. The search was limited to English-language papers only. An identical search was conducted in Google Scholar to identify any relevant tools or papers in the grey literature. Additional articles were identified by searching the reference lists of the articles found. No age criterion was applied.

Two authors (K.M., L.B.) sorted the articles for relevance and where disagreement arose a third author (M.M.) provided input. Where controversy remained, the relevance of the article was discussed and a final decision made regarding inclusion. Articles that referred to the development of a new tool to measure dairy and/or Ca intake in Western populations and did or did not test for validity or reliability were deemed relevant. Articles that (i) measured dairy and/or Ca intake in non-Western countries (due to differences in the major food sources of Ca), (ii) utilised existing tools but did not test these for validity or reliability in the study sample, (iii) were not in English, (iv) utilised 24 h recalls, food records or diet histories to measure dietary intake, (v) did not assess dairy or Ca intake or (vi) were published abstracts only were considered to be irrelevant for the purposes of the current review. From the initial 121 articles seventy-three were excluded, leaving forty-eight relevant articles for examination. Of these forty-eight, eighteen were categorised as relevant for children/adolescents.

Tools described in the articles were classified as (i) dairy assessment tools that assess the quantity or frequency of dairy food intake or (ii) Ca assessment tools that estimate absolute Ca intake or categorise intake into specific levels, e.g. ≥800 mg Ca/d or <800 mg Ca/d. Some tools assessed intake of dairy foods or dairy foods and other Ca-containing foods and quantified or classified Ca intake and quantified dairy food intake in terms of

quantity or frequency. These tools were considered to be both Ca and dairy assessment tools.

When assessing reliability and validity of tools a sample size of at least 100 subjects was considered acceptable[10], tests of association (correlation coefficients) were considered weak statistical analysis, whereas tests that measured agreement (Bland–Altman or kappa coefficient, κ) and/or sensitivity and specificity were considered to provide strong systematic analysis[11]. A mean difference of 100 mg (representing about 10 % of recommended daily intake or one-third of a serving of dairy products) was considered clinically significant. Further, a value κ >0·5 was considered moderate agreement, κ >0·7 good agreement and κ >0·8 very good agreement[11].

Results

Eighteen articles reporting on seventeen tools that had been used in those aged less than 18 years were identified (Table 1). For two tools there were two articles reporting on each[12–15], and another article reported on two other tools[16]. Only two of the tools assessed dairy intake, ten assessed Ca intake and five assessed both. Details of each of the tools are provided in Table 1. The tools were used in a range of population groups of differing age, gender and race and with the exception of the study by Taylor *et al.*[17], all subjects could be considered relatively healthy participants.

Tool characteristics

All tools used an FFQ, with varying response options covering a variable period. Ten were quantitative allowing an estimate of milligrams of Ca, six were semi-quantitative and one was qualitative. The quantitative tools allowed varying serving sizes, the semi-quantitative tools provided a standard serving size and the qualitative tool asked for frequency of consumption of each food only. In terms of food coverage, five tools included dairy products and other foods that make an important contribution to Ca intake. Two tools included Ca-containing foods plus those foods that may displace dairy or other high-Ca foods[18,19] and another included 'dummy' foods to mask the intent of the tool[20]. Three tools were designed to assess several nutrients and the food coverage reflected this[16,21,22]. The remaining six tools were general FFQ that were developed and tested for their ability to assess dairy and/or Ca intake.

Visual aids to assist participants to identify and quantify foods were provided with five tools. Six tools took less than 15 min to complete, indicating a level of user-friendliness, and one tool took 30 min[23]. Information regarding average completion times was not reported for ten tools; however, estimates from similar tools indicated that four of these could be completed in <15 min while the remaining six tools would likely require longer than this. Most tools required computer analysis or professional

Table 1 Summary of studies describing dairy and/or calcium assessment tools utilised in child/adolescent populations and their key features

Author (year), country	Tool*; tool type; population age; time span assessed	Food coverage	Time required to complete	Application
Huybrechts et al. (2011)[14,15], Europe	D:43; Q-FFQ; 2–10 years; typical week of previous month	Total diet	Unknown; likely <15 min	
Vereecken and Maes (2003)[22], Belgium	D:15; SQ-FFQ; 11–14 years; usual intake	Covers important food sources of Ca and fibre, and less healthy food items commonly consumed by youth: fruits, vegetables, cereals, bread, milk, cheese, other milk products, crisps, chips, sweets, soft drinks, alcoholic beverages	Unknown; likely <15 min	Not Ca/D-specific. Self-administered. Short and fast to complete. Appears computer calculated
Bellu et al. (1995)[21], Italy	Ca/D:116; Q-FFQ; 7–10 years; previous 6 months	Total diet: foods which account for up to 90 % of total intake of energy, lipids, carbohydrate, protein, cholesterol, Ca, Fe, vitamins (A, thiamin, riboflavin, B_6, C, E), soluble and non-soluble fibre	Unknown; likely ≥15 min	Not Ca/D-specific. Self-completed with parental assistance. Lengthy. Computer calculated
Marshall et al. (2008)[16], USA	Ca/D:22; SQ-FFQ; 9·0 (SD 0·7) years; previous week	Foods containing Ca, vitamin D, alcohol, caffeine, fluoride	Unknown; likely <15 min	Not Ca/D-specific. Completed with parent assistance. Computer calculated
	Ca/D:75; Q-FFQ; 8·3 (SD 0·3) years; previous week	Total diet	Not reported; approx 25 min as per http://www.nutritionquest.com/	Not Ca/D-specific. Completed with parent assistance. Computer calculated
Rockett et al. (1995)[26], USA	Ca/D:151; SQ-FFQ; 9–18 years; previous 12 months	Total diet	Unknown; likely ≥15 min	Not Ca/D-specific. Self-administered. Lengthy. Computer calculated
Vereecken et al. (2010)[29], Belgium	Ca/D:unknown (77 food groups); Q-FFQ; 2·6–5·5 years; previous 3 months	Total diet	Unknown; likely ≥15 min	Not Ca/D-specific. Lengthy. Parent-completed. Computer calculated
Bertoli et al. (2005)[23], Italy	Ca:136; Q-FFQ; 6–10 years and 16–20 years; unknown	Total diet	30 min (reported)	Not Ca/D-specific. Dietitian-administered with/without parent assistance. Computer calculated
Harnack et al. (2006)[27], USA	Ca:10; Q-FFQ; 11–14 years; previous month	Milk (dairy and non-dairy); Ca-fortified juice; yoghurt; cold cereal; pizza; macaroni and cheese; bread, toast or dinner rolls; Mexican foods; cheese	<5 min (reported)	Ca-specific. Self-administered. Short and fast to complete. Computer calculated
Huybrechts et al. (2006)[25], Belgium	Ca:47; SQ-FFQ; 2–6 years; previous 12 months	High-moderate Ca-containing foods: water; fruit juice; soya and other beverages; soup; milk; cheese; yoghurt; milk-based desserts; soya-based desserts; nuts and seeds; olives; fruit; chocolate; snack foods; breakfast cereals; breads and bread products; spreads; fish; meat; eggs; vegetarian products; potatoes; vegetables	15–20 min (estimated from experience)	Ca-specific. Parent-completed. Appears computer calculated

Table 1 Continued

Author (year), country	Tool*; tool type; population age; time span assessed	Food coverage	Time required to complete	Application
Jensen et al. (2004)[18], USA	Ca:80; SQ-FFQ; 10–18 years; previous month	Ca-containing foods plus drinks that displace consumption of milk and other high-Ca foods. Covers beverages; dairy products; combination foods; vegetables, grains and nuts; seafood; other foods	Unknown; likely ≥15 min	Ca-specific. Self-administered. Computer calculated
Magkos et al. (2006)[12,13], Greece	Ca:30; SQ-FFQ; 11·9 (SD 1·2) years; previous 12 months	Dairy (milk, yoghurt, cheese), pies, cereal products, nuts, vegetable products, legumes, fish products, eggs, ice cream and chocolate	~5 min (reported)	Ca-specific. Self-administered. Short and fast to complete. Suitable for both genders and all age groups. Requires computer or professional to calculate Ca intake
Marcotte et al. (2008)[20], USA	Ca:19; SQ-FFQ; 6–10 years; previous 24 h	Ca-containing foods plus 'dummy foods' to mask purpose of tool. Includes juice, milk, yoghurt, cheese, soda, soup, cereal, snack bar, waffles/pancakes, frozen desserts, pudding/ custard, tofu, rice and beans, peanut butter and jelly, hotdog with roll, broccoli/kale/bok choy, hamburger on bun, pizza, macaroni and cheese	Unknown; likely <15 min	Ca-specific, plus dummy foods. Interviewer-administered. Hand calculated
Taylor et al. (2009)[17], USA	Ca:40; SQ-FFQ; 12–18 years, females with anorexia nervosa; unknown	Ca-containing and vitamin D-containing foods. Includes dairy products, mixed dishes containing dairy, green vegetables, Ca-fortified juice, bread and cereal products/dishes, legumes, tofu, eggs, seafood, cake, almonds, chocolate, oral nutrition supplements, vitamin supplements	10–15 min (estimated from experience)	Not Ca/D-specific. Moderate length. Dietitian-administered. Hand calculated
Wong et al. (2008)[19], USA	Ca:80; SQ-FFQ; 11–18 years; previous month	Ca-containing foods, drinks that displace consumption of milk and other high-Ca foods. Covers beverages, dairy products, combination foods, vegetables, grains and nuts, seafood, other foods	Unknown; likely ≥15 min	Ca-specific. Self-administered. Requires computer access and computer literacy
Yang et al. (2010)[24], USA	Ca:15; SQ-FFQ; 9–20 years; previous month	Ca-rich foods: milk, milk-based drinks, yoghurt, frozen yoghurt or ice cream, cheese, cold cereal, chocolate bar, mixed dishes including diary, tofu	~5 min (reported)	Ca-specific. Self-administered and self-calculated. Short and fast to complete. Converts Ca intake to a Ca score ('excellent', 'good', 'fair' or 'poor')
Zemel et al. (2010)[28], USA	Ca:41; Q-FFQ; 7–10 years; previous 4 weeks	Foods contributing >50 mg Ca/d from dietary intake survey in target population. Includes beverages, dairy products, combination foods and other Ca sources, e.g. kale, milk chocolate bars, Ca-fortified items, nutritional supplements	~15 min to complete FFQ (reported). 15–30 min to analyse (reported)	Ca-specific. Dietitian-administered with parental assistance. Computer calculated

D, dairy; Q-FFQ, quantitative FFQ; SQ-FFQ, semi-quantitative FFQ.
*Purpose: number of items in the tool.

assistance to determine total daily Ca intake and/or adequacy. A few tools were able to provide an immediate indication of daily Ca intake, through a computer-based FFQ which automatically calculates intake for the participant[19], or via manual calculation with or without reference to the recommended dietary intake[17,20], or via a simple scoring system to grade adequacy of intake[24].

Tool properties

All tools had one or more properties tested. Only ten papers reported test–retest reliability and these were for two dairy tools, one dairy/Ca tool and seven Ca tools (Table 2). The statistical analyses varied with correlation (Pearson, Spearman or intra-class) being the most frequently used test and six studies reporting mean values for each administration, with only Lanfer et al.[15] not reporting standard deviations or conducting t-test analysis. Three studies reported κ values[15,22,25] and two of these studies also reported classification agreement[22,25]. The tools were mostly used in children aged 10 years and older, with eight of the studies having a sample size over 100 and the other two less than fifty. The period between the two administrations of the tool varied from a minimum of 1 h[20] to 1 year[26], with most being 1 to 4 weeks. Rockett et al. reported reliability for only one dairy item for which the correlation was modest (0·56)[26]. The more comprehensive results of the study by Veerecken and Maes assessing four dairy items in two age groups demonstrated only moderate repeatability for most of the items with the younger age group tending to perform better than the older group[22]. Similarly, the five dairy items in the Child Eating Habits Questionnaire assessed by Lanfer et al. also showed moderate repeatability[15].

Huybrechts et al. reported a weighted κ value of 0·60 representing only 'good' agreement for their tool but a high correlation (0·80), a non-significant mean difference of only 24 mg, and 57 % correctly classified and no gross misclassification[25]. Only two other studies[18,19] reported correlation values as high as Huybrechts et al.[25] and both studies explored test–retest reliability by ethnicity (Asian, Hispanic, non-Hispanic White) and by various age groups. While correlation was highest for 14/15- to 18-year-olds compared with 10/11- to 13/14-year-olds, there were no clear ethnic group or gender differences. One of these two studies[19] reported a statistically significant mean difference of approximately 200 mg Ca between repeat administrations which is clinically unacceptable. Similarly, Harnack et al. demonstrated a clinically unacceptable and statistically significant difference in mean Ca, on repeat administration of their tool, of ~100 mg despite an acceptable intra-class correlation of 0·66 to 0·79[27].

Two other studies[24,28] reported correlations (0·74, 0·76) near that of Huybrechts et al.[25] and Yang et al.[24] also reported a non-significant mean difference on repeat administrations. Lower correlations were reported by Rockett et al. (0·58), who also reported a significant and unacceptable mean difference of 150 mg[26], and Marcotte et al. (0·49–0·67), who reapplied the tool in many cases only 1 h after the first administration, and hence the true correlation for a more appropriate repeat administration of at least 1 week remains unknown[20].

Fifteen studies reported relative validity, three providing data on dairy food intake using four tools[16,22,29] and fourteen reporting Ca intake using fourteen tools (Table 3). A range of reference methods were used across the fifteen papers. Three studies reported using a 7 d weighed food record[21,23,28] and five studies reported using an estimated food record of 3 to 7 d[16,17,25,29], one of which was web-based[29]. Three studies used a single 24 h recall[12,13,20] and four[18,19,27,30] used multiple 24 h recalls. Huybrechts et al. assessed validity v. urinary Ca excretion[14].

An array of statistical tests were performed and included correlation, comparison of mean values, Bland–Altman plots, agreement using κ, cross-classification, and assessment of sensitivity and specificity (Table 3).

Four tools, reported in three papers, estimated various dairy products with both mean values and correlations for the tool and reference method presented[16,22,29]. With the exception of the seventy-five-item Block Kids' Food Questionnaire reported by Marshall et al., all studies showed significant differences for milk intake between the test and reference tool. Similarly there were significant differences for most other dairy products. While correlations for milk questions were good to high, they were poor for other dairy foods. Values of κ and cross-classification results generally supported the poor performance of the tools.

All studies using tools to estimate Ca intake reported results of a correlation test between the estimate for the tool and the reference method. Few results were greater than 0·5, suggesting moderate association at best. Thus a discussion of the values reported for correlation coefficients (Pearson, Spearman, intra-class) will not be provided in the present review as the meaningfulness of these analyses is questionable. However, the values are reported in Table 3 to allow the reader to consider these alongside other findings.

All studies also reported mean Ca intake estimated by the tool and the reference method. Eight papers reporting on seven tools showed significant differences in the overall sample or in subgroups[12,13,18,23,25,27–29]. In addition, Marshall et al. reported significant mean differences for the twenty-two-item targeted nutrient questionnaire but not for the seventy-five-item Block Kids' Food Questionnaire[16]. Four other authors found no statistically significant differences in comparison with the reference method[17,20,21,30]. For these studies the mean differences ranged from values that would not be considered clinically meaningful, <10 mg[16] and <20 mg[21], to values that are approaching concern: ~66 mg[30],

Table 2 Details of reliability testing of tools that assess calcium and/or dairy intake in children and adolescents

Author (year), country	Tool*	Population	Period between administrations	Paired *t* test	Correlation: Pearson, transformed Pearson, Spearman, intra-class correlation	Cross-classification, κ
					Statistical tests	
Vereecken and Maes (2003)[22], Belgium	D:15	*n* 207, 11–12 years; *n* 560, 13–14 years	11–12 years: 7–15 d 13–14 years: 6–10 d		S: 4 dairy items 11–12 years: *r* = 0·59–0·82 13–14 years: *r* = 0·57–0·71	CC: 11–12 years, 4 dairy items: 39–68 % agreement, 74–88 % same/adjacent CC: 13–14 years, 3 dairy items: 37–62 % agreement, 70–80 % same/adjacent κ: 11–12 years, 4 dairy items: 0·44–0·70; 13–14 years, 3 dairy items: 0·59–0·61
Lanfer *et al.* (2011)[15], Europe	D:43	*n* 258, 2–9 years	0–354 d		S: 5 dairy items, *r* = 0·60–0·76	κ: 5 dairy items: 0·50–0·68
Rockett *et al.* (1995)[26], USA	Ca/D:151	*n* 179, 9–18 years	12 months	1241 (SD 637) mg *v.* 1092 (SD 540) mg, *P* < 0·05	P: *r* = 0·58	
Harnack *et al.* (2006)[27], USA	Ca:10	*n* 248, 11–14 years, ethnically diverse	7 d	856 (SD 570) mg *v.* 752 (SD 551) mg, *P* < 0·001	ICC: 0·66–0·79	
Huybrects *et al.* (2006)[25], Belgium	Ca:47	*n* 509, 2·5–6·5 years	5 weeks	774 (SD 252) mg *v.* 751 (SD 255) mg, *P* = 0·26	P: *r* = 0·80	CC: 56·7 % correct, 0 % grossly misclassified κ: 0·60 (95 % CI 0·49, 0·71)
Jensen *et al.* (2004)[18], USA	Ca:80	*n* 162, 10–18 years, Asian, Hispanic and white youth	4 weeks		P: *r* = 0·46–0·81 TP: *r* = 0·48–0·77 S: *r* = 0·40–0·77	
Marcotte *et al.* (2008)[20], USA	Ca:19	*n* 42, 7·1–8·9 years	Min 1 h		S: *r* = 0·49–0·67, *P* < 0·05	
Wong *et al.* (2008)[19], USA	Ca:80	*n* 161, 11–18 years	4 weeks	1210 (SD 626) mg *v.* 1045 (SD 678) mg, *P* < 0·05	P: *r* = 0·49–0·80 TP: *r* = 0·59–0·82 S: *r* = 0·61–0·80	
Zemel *et al.* (2010)[28], USA	Ca:41	*n* 139, 7–10 years	Max 3·6 months (median 32 d)		P: *r* = 0·74	
Yang *et al.* (2010)[24], USA	Ca:15	*n* 41, 9–17 years	7 d	*t* = −0·274, *P* = 0·786	P: *r* = 0·757, *P* < 0·0001	

D, dairy; *n*, sample size; min, minimum; max, maximum; S, Spearman's correlation; P, Pearson's correlation; ICC, intra-class correlation; TP, transformed Pearson's correlation; CC, cross-classification; κ, kappa coefficient.
*Purpose: number of items in the tool.

Table 3 Details of validity testing of tools that assess calcium and/or dairy intake in children and adolescents

Author (year), country	Tool*; reference tool; population	Statistical tests			
		Correlation: Pearson or Spearman	Paired t test or Mann–Whitney U test or Wilcoxon signed-rank test	Cross-classification or intra-class correlation	Sensitivity/specificity or Bland–Altman or κ
Vereecken and Maes (2003)[22], Belgium	D:15 7 d food diary n 207, 11–12 years; n 560, 13–14 years	Semi-skimmed milk: $r = 0.65$ Whole-fat milk: $r = 0.64$ Cheese: $r = 0.49$ Other milk products: $r = 0.41$	Diary v. FFQ (times/week): Semi-skimmed milk: 2.9 v. 4, $P < 0.05$ Whole-fat milk: 1.42 v. 2, $P < 0.05$ Cheese: 2.36 v. 2.55, $P > 0.05$ Other milk products: 3.24 v. 3.83, $P < 0.05$	CC: Agreement: Semi-skimmed milk = 62 % Whole-fat milk = 72 % Cheese = 53 % Other milk products = 49 % Gross misclassification: Semi-skimmed milk = 14 % Whole-fat milk = 7 % Cheese = 7 % Other milk products = 3 %	
Bellu et al. (1995)[21], Italy	Ca/D:116 7 d weighed food diary: n 88, 7–10 years 24-h recall: n 323, 7–10 years	Diary v. FFQ, $r = 0.322$	Recall v. FFQ: G: 441 mg v. 445 mg, $P = 0.80$ B: 453 mg v. 435 mg, $P = 0.26$		
Marshall et al. (2008)[16], USA	Ca/D:22 and Ca/D:75 3 d food diary n 223 and n 129, mean 8–9 years	Ca/D:22: $r = 0.462$ Ca/D:75: $r = 0.515$	Diary v. Ca/D:22: Ca: 910 mg v. 733 mg, $P < 0.05$ Milk: 338 g v. 421 g, $P < 0.05$ Diary v. Ca/D:75: Ca: 906 mg v. 915 mg, $P > 0.05$ Milk: 427 g v. 417 g, $P > 0.05$	CC: Correct for Ca: Ca/D:22 = 46.1 % Ca/D:75 = 34.9 % Correct for milk: Ca/D:22 = 44.8 % Ca/D:75 = 46.0 %	κ: For Ca: Ca/D:22 = 0.38 Ca/D:75 = 0.28 For milk: Ca/D:22 = 0.42 Ca/D:75 = 0.43
Rockett et al. (1997)[30], USA	Ca/D:151 Mean of three 24 h recalls n 261, 9–18 years	$r = 0.61$	Recall v. FFQ: 1093 (SD 454) mg v. 1159 (SD 417) mg		
Vereecken et al. (2010)[29], Belgium	Ca/D:77 3 d web-based food record n 216, 2.6–5.5 years	Ca: $r = 0.59$ Milk: $r = 0.76$ Sugared milk beverages: $r = 0.68$ Milk-based desserts and cream: $r = 0.49$ Cheese: $r = 0.35$	Web record v. FFQ: Ca: 796 mg v. 724 mg, $P < 0.05$ Milk: 202 g v. 140 g, $P < 0.05$ Sugared milk beverages: 148 g v. 129 g, $P < 0.05$ Milk-based desserts and cream: 53 g v. 81 g, $P < 0.05$ Cheese: 11 g v. 8 g, $P < 0.05$		BA: Mean bias: <50 mg LOA: −400, 500 mg

Table 3 *Continued*

Author (year), country	Tool*; reference tool; population	Statistical tests			
		Correlation: Pearson or Spearman	Paired *t* test or Mann–Whitney *U* test or Wilcoxon signed-rank test	Cross-classification or intra-class correlation	Sensitivity/specificity or Bland–Altman or κ

Bertoli *et al.* (2005)[23], Italy	Ca:136
	7 d weighed food diary
	n 18, 6–10 years; *n* 19, 16–20 years
		6–10 years: *r* = 0·6
		16–20 years: *r* = 0·5
			Diary *v.* FFQ:
			  6–10 years: 691 mg *v.* 1039 mg, *P* < 0·05
			  16–20 years: 726 mg *v.* 1073 mg, *P* < 0·05
				CC:
				  Correct: overall 62 %; tertiles 75/50/61 %
				  Gross misclassification: tertiles 0/0/7 %
					BA:
					  Mean bias: ~250 mg
					  LOA: ~−250, ~750 mg
					κ: 0·34

Harnack *et al.* (2006)[27], USA	Ca:10
	Mean of three 24 h recalls
	n 248, 11–14 years, ethnically diverse
			Recall *v.* FFQ:
			  G: 907 mg *v.* 826 mg, *P* = 0·07
			  B: 1125 mg *v.* 903 mg, *P* < 0·05
			  11–12 years: 1008 mg *v.* 856 mg, *P* < 0·05
			  13–14 years: 985 mg *v.* 856 mg, *P* < 0·05
			  Total: 993 mg *v.* 856 mg, *P* < 0·05
				ICC:
				  G: 0·45
				  B: 0·40
				  11–12 years: 0·59
				  13–14 years: 0·33
				  Total: 0·43	

Huybrechts *et al.* (2006)[25], Belgium	Ca:47
	3 d food diary
	n 509, 2·5–6·5 years
		r = 0·52	Diary *v.* FFQ:
			  838 mg *v.* 777 mg, *P* < 0·05
				CC:
				  Correctly classified: 42 %
				  Within 1 quartile: 41 %
				  Gross misclassification: 2·4 %
					SE: 62 %, SP: 77 %
					BA:
					  Mean bias: 60·9 mg,
					  LOA: ~−500, ~600 mg
					κ: 0·38

Jensen *et al.* (2004)[18], USA	Ca:80
	Mean of two 24 h recalls
	n 162, 10–18 years, A, H, W
		Pearson's:
		  G: *r* = 0·31
		  B: *r* = 0·51
		  10–13 years: *r* = 0·32
		  14–18 years: *r* = 0·53
		  A: *r* = 0·48
		  H: *r* = 0·16
		  W: *r* = 0·49
		Spearman's:
		  G: *r* = 0·37
		  B: *r* = 0·55
		  10–13 years: *r* = 0·35
		  14–18 years: *r* = 0·49
		  A: *r* = 0·53
		  H: *r* = 0·21
		  W: *r* = 0·43
			Recall *v.* FFQ:
			  G: 974 mg *v.* 1138 mg, *P* < 0·05
			  B: 1251 mg *v.* 1176 mg, *P* > 0·05
			  10–13 years and 14–18 years, *P* > 0·05
			  H: 903 mg *v.* 1185 mg, *P* < 0·05
			  A and W, *P* > 0·05		

Table 3 *Continued*

		Statistical tests			
Author (year), country	Tool*; reference tool; population	Correlation: Pearson or Spearman	Paired *t* test or Mann–Whitney *U* test or Wilcoxon signed-rank test	Cross-classification or intra-class correlation	Sensitivity/specificity or Bland–Altman or κ

Magkos *et al.* (2006)[12], Greece	Ca:30
	24 h recall
	n 351, 10–15 years
		G: *r* = 0·625
		B: *r* = 0·638
			Recall *v.* FFQ:
			  G: 946 mg *v.* 885 mg, *P* < 0·05
			  B: 995 mg *v.* 906 mg, *P* < 0·05
				CC:
				  Correct: G = 41 %, B = 45 %
				  Within 1 quartile: G = 46 %, B = 39·5 %
				  Gross misclassification:
				    G = 2·6 %, B = 1·2 %
					SE: G = 74·1 %, B = 70·0 %
					SP: G = 74·1 %, B = 72·5 %
					BA:
					  Mean bias: G = 62 mg, B = 89 mg
					  LOA: G = −747, 624 mg, B = −762, 585 mg
					κ: G = 0·558, B = 0·566

Magkos *et al.* (2006)[13], Greece	Ca:30
	24 h recall
	n 351, 10–15 years
		r = 0·639	Recall *v.* FFQ:
			  861 mg *v.* 728 mg, *P* < 0·05
				CC:
				  Correct: 41·5 %
				  Within 1 quartile: 41·1 %
				  Gross misclassification: 1·7 %
					SE: 82·8 %, SP: 54·9 %
					BA:
					  Mean bias: 133 mg
					  LOA: −799, 533 mg

Marcotte *et al.* (2008)[20], USA	Ca:19
	24 h recall
	n 42, 7·1–8·9 years
		G: *r* = 0·65
		B: *r* = 0·33
		Total: *r* = 0·28
			Recall *v.* checklist:
			  G: 931 mg *v.* 898 mg
			  B: 1217 mg *v.* 983 mg
			  Total: 1053 mg *v.* 935 mg, *P* > 0·05
					BA:
					  Mean bias: G = ~250 mg, B = ~125 mg
					  LOA: G = ~−1300, ~1800 mg, B = ~−800, ~1100 mg

Taylor *et al.* (2009)[17], USA	Ca:40
	4 d food diary
	n 36, 12–18 years, females, AN; *n* 39, 12–18 years, females, C
		C: *r* = 0·45
		AN: *r* = 0·65
		Total: *r* = 0·60
			Diary *v.* FFQ:
			  C: 908 mg *v.* 1095 mg, *P* > 0·05
			  AN: 1485 mg *v.* 1429 mg, *P* > 0·05
			  Total: 1185 mg *v.* 1255 mg, *P* > 0·05
				CC:
				  Correct: C = 36 %, AN = 53 %, total: 49 %
				  Gross misclassification:
				    C = 5 %, AN = 0 %	

Wong *et al.* (2008)[19], USA	Ca:80
	Mean of two 24 h recalls
	n 161, 11–18 years, A, H, W
		Pearson's:
		  G: *r* = 0·46
		  B: *r* = 0·35
		    11–14 years: *r* = 0·44
		    15–18 years: *r* = 0·46
		  A: *r* = 0·54
		  H: *r* = 0·39
		  W: *r* = 0·46
		  Total: *r* = 0·42
		Spearman's:
		  G: *r* = 0·47
		  B: *r* = 0·48
		    11–14 years: *r* = 0·47
		    15–18 years: *r* = 0·54
		  A: *r* = 0·57
		  H: *r* = 0·50
		  W: *r* = 0·45
		  Total: *r* = 0·51
			Recall *v.* FFQ:
			  1172 mg *v.* 1045 mg
			  (*P* not reported)		

Table 3 Continued

Author (year), country	Tool*; reference tool; population	Statistical tests			
		Correlation: Pearson or Spearman	Paired t test or Mann–Whitney U test or Wilcoxon signed-rank test	Cross-classification or intra-class correlation	Sensitivity/specificity or Bland–Altman or κ
Zemel et al. (2010)[28], USA	Ca:41 7 d weighed food diary n 139, 7–10 years, AA and NAA	G: r=0.59 B: r=0.64 7–8 years: r=0.43 9–10 years: r=0.67 AA: r=0.35, NAA: r=0.66 Total: r=0.61	Diary v. FFQ: G: 838 mg v. 1190 mg, P<0.05 B: 950 mg v. 1194 mg, P<0.05 7–8 years: 776 mg v. 1099 mg, P<0.05 9–10 years: 937 mg v. 1229 mg, P<0.05 AA: 822 mg v. 1058 mg, P>0.05 NAA: 901 mg v. 1249 mg, P<0.05 Total: 885 mg v. 1192 mg, P<0.05	CC: Correctly classified: 39% Within 1 quintile: 35%	

D, dairy; n, sample size; A, Asian; H, Hispanic; W, white; AN, anorexia nervosa; C, control; AA, African American; NAA, non-African American; G, girls; B, boys; CC, cross-classification; ICC, intra-class correlation; κ, kappa coefficient; BA, Bland-Altman; LOA, limits of agreement; SE, sensitivity; SP, specificity.
*Purpose: number of items in the tool.

~70 mg[17] and ~120 mg[20]. While no P value was provided, the mean difference reported by Wong et al. was 127 mg[19].

Only three studies reported sensitivity and specificity analysis[12,13,25] (Table 3). Sensitivity values ranged from 62%[25] to 82·8%[13] and specificity values ranged from 54·9%[13] to 77%[25]. A similar concept to sensitivity and specificity is the calculation of cross-classification statistics, a model applied by far more studies (n 8)[12,13,16–18,23,25,28]. While one study[23] did achieve levels of correct classification similar to those that reported sensitivity and specificity[12,13,25], none of these additional studies achieved a smaller level of gross misclassification than those studies (5–26% v. 1–2·6%).

Six studies used Bland–Altman plots to illustrate the level of agreement[12,13,20,23,25,29]. The plots identify the mean bias and limits of agreement between methods and identify whether a tool is valid for assessment of Ca intake at the individual and population levels[31]. The mean bias between the tool and reference method ranged from ~50 mg[29] to 250 mg[20,23]. The difference between the lower and upper limit of agreement varied widely between studies but all were greater than 900 mg, indicating that none of the tools evaluated would be suitable for use at an individual level.

Discussion

The current review identified eighteen papers that reported on two tools that assessed dairy intake, ten that assessed Ca intake and five that assessed both dairy and Ca. Overall assessment of tool properties indicated a general lack of evidence to support use of any of the tools. The most common limitation of the testing of tool properties was the high reliance on correlation, which assesses association but not agreement. Ideally, validity is assessed using a measure of agreement such as Bland–Altman analysis[8] and a measure of sensitivity and specificity (the ability of a test to correctly identify true positives and true negatives as in the case of identifying adequate intake).

In addition, when assessing validity it is important to determine a clinically meaningful level of significance as opposed to statistical significance. None of the papers defined a level of clinical significance at which the results were meaningful in terms of dietary adequacy. This lack of recognition between statistically and clinically significant results limits conclusions relevant to clinical practice. We applied a 100 mg Ca cut-off when assessing studies.

Tests of reliability were reported for ten tools only and of these only four studies reported mean difference for the repeat administration (a fifth reported the P value only) and only two reported cross-classification. Based on these results and the overall lack of tests, only one tool assessing Ca[25] and one tool assessing dairy[22] could be considered reliable.

Validation studies of the three tools assessing dairy intake suggest that these tools are not valid for estimating dairy intake. With respect to tools that assess Ca, validity results vary such that some should be considered with caution while others appear to have acceptable levels of agreement and/or sensitivity and specificity indicating adequate validity. The tools that appear to be best in levels of validity are those developed by Huybrechts et al.[25], Vereecken et al.[29] and Magkos et al.[12]. Each of these studies included a minimum of 216 children and reported less than 100 mg difference between the test tool and the reference method. It is important to note that the two Belgian studies were tested in children aged ~2 to 7 years[25,29], while a more suitable tool for older children might be the tool reported by Magkos et al. as it has been tested in children aged 10 to 15 years[12]. Importantly, only one of these tools was tested for reliability[25].

In addition to the results of testing of tool properties it is important to consider the quality of the study design used in providing these findings. Key study design criteria include level of evidence, potential sources of error and bias, and sample size. All studies had level III-2 level of evidence, as defined by the National Health and Medical Research Council evidence hierarchy for diagnostic accuracy[32], i.e. there was a comparison with a reference standard. They do not meet the criterion for higher evidence which requires an independent blinded comparison with a valid reference standard, thus introducing a risk of positive respondent bias.

Possible sources of error in the methodology are limitations associated with the reference methods used for comparison with the identified tools. As there is no biomarker for Ca status, only relative validity can be assessed, i.e. against another method, and thus there will be some inherent limitations. Sample size is important for reliability and validation studies with a sample size of 100 suggested as the minimum[10]. Most studies, and those potentially recommended, met this criterion.

Potential positive respondent bias was present across all of the studies as none were blinded; thus participants were aware of the study purpose and may have altered their responses with the aim of conforming to the recommendations regarding Ca or dairy intake, resulting in over-reporting of intake. Potential recall bias is always associated with an FFQ and the degree will depend on the cognitive ability of the subjects and the period to which the FFQ refers (e.g. the last week, the last year). With the exception of one study, samples appeared representative of the population indicating absence of recruitment bias.

The ability for a tool to detect change is an additional attribute and some argue this should be a third essential property of a tool[33]. Sensitivity to change describes a tool's ability to measure any degree of change and responsiveness describes a tool's ability to measure clinically important change. These parameters can be measured in a number of ways. The most frequently used measure is Cohen's effect size: the ratio of the mean difference (of the two measures) to the standard deviation at baseline. Whether or not a statistically significant change is detected depends on the sample size. However, if the sample is large a small change will be significant although such a change may not be clinically significant. Importantly, it is not appropriate to measure change within subjects if the reliability of the tool is less than 0·5[33]. None of the tools described/recommended specifically reported the tool's ability to detect change.

Based on the review of methods and results of the reliability and validity testing of identified tools, there is no recommended tool for assessing dairy intake in children but three for assessing Ca intake. Appropriate testing methods for validity and demonstrated adequate levels of sensitivity, specificity and/or agreement were reported for all these tools but only one was tested for reliability, which was shown to be moderate[25]. Interestingly, there were few common elements between these tools with the number, type and scope of questions and the target group varying. The tool by Magkos et al. was the shortest, containing thirty questions, assessing semi-quantitatively the intake of principally Ca-containing foods only in the previous 12 months and self-completed by children aged 10–14 years[12]. The tool by Huybrechts et al. was parent-completed for young children (2–6 years), contained forty-seven items most of which were high Ca-containing foods, assessed semi-quantitatively for the previous 12 months[25]. The tool reported by Vereecken et al. was also parent-completed for very young children but had seventy-seven items assessing quantitatively total diet in the previous 3 months[29]. In conclusion, there is an obvious need for development and rigorous testing of tools to assess Ca and/or dairy intake in children and adolescents.

Acknowledgements

References

1. Australian Bureau of Statistics (1998) *National Nutrition Survey: Nutrient Intakes and Physical Measurements, Australia, 1995. Catalogue* no. 4805·0. Canberra: ABS.
2. Heaney RP (2009) Dairy and bone health. *J Am Coll Nutr* **28**, Suppl. 1, 82S–90S.

3. Spence LA (2013) Shortfall in the consumption of dairy foods in Australian children's diets: addressing health implications and practical approaches. *Nutr Diet* **70**, 66–70.

4. Heaney RP (2000) Calcium, dairy products and osteoporosis. *J Am Coll Nutr* **19**, 2 Suppl., 83S–99S.

5. UK Department of Health (2011) *National Diet and Nutrition Survey: Headline Results from Years 1 and 2 (combined) of the Rolling Programme 2008/9 and 2009/10.* London: Department of Health.

6. US Department of Agriculture, Agricultural Research Service (2010) *What We Eat in America, NHANES 2007–2008, Nutrient Intakes from Food: Mean Amounts Consumed per Individual, by Gender and Age.* Washington, DC: US Department of Agriculture.

7. Baird DL, Syrette J, Hendrie GA *et al.* (2012) Dairy food intake of Australian children and adolescents 2–16 years of age: 2007 Australian National Children's Nutrition and Physical Activity Survey. *Public Health Nutr* **15**, 2060–2073.

8. Lang T & Secic M (2006) *How To Report Statistics in Medicine*, 2nd ed. USA: Sheridan Press.

9. Jones JM (2004) Validity of nutritional screening and assessment tools. *Nutrition* **20**, 312–317.

10. Willett W (1998) *Nutritional Epidemiology*, 2nd ed. New York: Oxford University Press.

11. Peat J, Mellis C, Williams K *et al.* (2001) *Health Science Research: A Handbook of Quantitative Methods.* Crows Nest, NSW: SAGE Publications.

12. Magkos F, Manios Y, Babaroutsi E *et al.* (2006) Development and validation of a food frequency questionnaire for assessing dietary calcium intake in the general population. *Osteoporos Int* **17**, 304–312.

13. Magkos F, Manios Y, Babaroutsi E *et al.* (2006) Differences in the quantitative and qualitative performance of a calcium-specific food frequency questionnaire across age and sex. *J Hum Nutr Diet* **19**, 331–342.

14. Huybrechts I, Börnhorst C, Pala V *et al.* (2011) Evaluation of the Children's Eating Habits Questionnaire used in the IDEFICS study by relating urinary calcium and potassium to milk consumption frequencies among European children. *Int J Obes (Lond)* **35**, Suppl. 1, S69–S78.

15. Lanfer A, Hebestreit A, Ahrens W *et al.* (2011) Reproducibility of food consumption frequencies derived from the Children's Eating Habits Questionnaire used in the IDEFICS study. *Int J Obes (Lond)* **35**, Suppl. 1, S61–S68.

16. Marshall TA, Eichenberger Gilmore JM, Broffitt B *et al.* (2008) Relative validity of the Iowa Fluoride Study targeted nutrient semi-quantitative questionnaire and the Block Kids' Food Questionnaire for estimating beverage, calcium, and vitamin D intakes by children. *J Am Diet Assoc* **108**, 465–472.

17. Taylor C, Lamparello B, Kruczek K *et al.* (2009) Validation of a food frequency questionnaire for determining calcium and vitamin D intake by adolescent girls with anorexia nervosa. *J Am Diet Assoc* **109**, 479–485.

18. Jensen JK, Gustafson D, Boushey CJ *et al.* (2004) Development of a food frequency questionnaire to estimate calcium intake of Asian, Hispanic, and white youth. *J Am Diet Assoc* **104**, 762–769.

19. Wong SS, Boushey CJ, Novotny R *et al.* (2008) Evaluation of a computerized food frequency questionnaire to estimate calcium intake of Asian, Hispanic, and non-Hispanic white youth. *J Am Diet Assoc* **108**, 539–543.

20. Marcotte L, Hennessy E, Dwyer J *et al.* (2008) Validity and reliability of a calcium checklist in early elementary-school children. *Public Health Nutr* **11**, 57–64.

21. Bellu R, Riva E, Ortisi MT *et al.* (1995) Calcium intakes in a sample of 35,000 Italian schoolchildren. *J Int Med Res* **23**, 191–199.

22. Vereecken CA & Maes L (2003) A Belgian study on the reliability and relative validity of the Health Behaviour in School-Aged Children food-frequency questionnaire. *Public Health Nutr* **6**, 581–588.

23. Bertoli S, Petroni ML, Pagliato E *et al.* (2005) Validation of food frequency questionnaire for assessing dietary macronutrients and calcium intake in Italian children and adolescents. *J Pediatr Gastroenterol Nutr* **40**, 555–560.

24. Yang YJ, Martin BR & Boushey CJ (2010) Development and evaluation of a brief calcium assessment tool for adolescents. *J Am Diet Assoc* **110**, 111–115.

25. Huybrechts I, De Bacquer D, Matthys C *et al.* (2006) Validity and reproducibility of a semi-quantitative food-frequency questionnaire for estimating calcium intake in Belgian preschool children. *Br J Nutr* **95**, 802–816.

26. Rockett HRH, Wolf AM & Colditz GA (1995) Development and reproducibility of a food frequency questionnaire to assess diets of older children and adolescents. *J Am Diet Assoc* **95**, 336–340.

27. Harnack LJ, Lytle LA, Story M *et al.* (2006) Reliability and validity of a brief questionnaire to assess calcium intake of middle-school-aged children. *J Am Diet Assoc* **106**, 1790–1795.

28. Zemel BS, Carey LB, Paulhamus DR *et al.* (2010) Quantifying calcium intake in school age children: development and validation of the Calcium Counts!© food frequency questionnaire. *Am J Hum Biol* **22**, 180–186.

29. Vereecken C, Covents M & Maes L (2010) Comparison of a food frequency questionnaire with an online dietary assessment tool for assessing preschool children's dietary intake. *J Hum Nutr Diet* **23**, 502–510.

30. Rockett HR, Breitenbach M, Frazier AL *et al.* (1997) Validation of a youth/adolescent food frequency questionnaire. *Prev Med* **26**, 808–816.

31. Bland JM & Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1**, 307–310.

32. National Health and Medical Research Council (2009) *NHMRC Levels of Evidence and Grades for Recommendations for Developers of Guidelines.* Canberra: NHMRC.

33. Streiner DL & Norman GR (2008) *Health Measurement Scales: A Practical Guide to Their Development and Use.* Oxford: Oxford University Press.