Published in final edited form as:

J Am Diet Assoc. 2004 September ; 104(9): 1385–1392.

Assessment of Interobserver Reliability in Nutrition Studies that Use Direct Observation of School Meals

MICHELLE L. BAGLIO, RD

M. L. Baglio, Georgia Prevention Institute, Department of Pediatrics, Medical College of Georgia, Augusta

SUZANNE DOMEL BAXTER, PhD, RD, FADA

S. D. Baxter is a research professor, Center for Research in Nutrition and Health Disparities, Department of Epidemiology and Biostatistics, University of South Carolina, Columbia

CAROLINE H. GUINN, RD

C. H. Guinn is a research dietitian, Center for Research in Nutrition and Health Disparities, Department of Epidemiology and Biostatistics, University of South Carolina, Columbia

WILLIAM O. THOMPSON, PhD

W. O. Thompson is director and professor emeritus, Office of Biostatistics and Bioinformatics, Medical College of Georgia, Augusta.

NICOLE M. SHAFFER, RD

N. M. Shaffer, Georgia Prevention Institute, Department of Pediatrics, Medical College of Georgia, Augusta

FRANCESCA H. A. FRYE, MBA, RD

F. H. A. Frye, Georgia Prevention Institute, Department of Pediatrics, Medical College of Georgia, Augusta

Abstract

This article (a) provides a general review of interobserver reliability (IOR) and (b) describes our method for assessing IOR for items and amounts consumed during school meals for a series of studies regarding the accuracy of fourth-grade children's dietary recalls validated with direct observation of school meals. A widely used validation method for dietary assessment is direct observation of meals. Although many studies utilize several people to conduct direct observations, few published studies indicate whether IOR was assessed. Assessment of IOR is necessary to determine that the information collected does not depend on who conducted the observation. Two strengths of our method for assessing IOR are that IOR was assessed regularly throughout the data collection period and that IOR was assessed for foods at the item and amount level instead of at the nutrient level. Adequate agreement among observers is essential to the reasoning behind using observation as a validation tool. Readers are encouraged to question the results of studies that fail to mention and/or to include the results for assessment of IOR when multiple people have conducted observations.

Address correspondence to: Suzanne Domel Baxter, PhD, RD, FADA, Arnold School of Public Health, Department of Epidemiology and Biostatistics, Center for Research in Nutrition and Health Disparities, University of South Carolina, 220 Stoneridge Dr, Suite 103, Columbia, SC 29210. E-mail: sbaxter@gwm.sc.edu.

This research was supported by R01 grant HL 63189 from the National Heart, Lung, and Blood Institute of the National Institutes of Health. Suzanne Domel Baxter, PhD, RD, FADA was the principal investigator. The authors express appreciation to the children, faculty, and staff of Blythe, Goshen, Gracewood, Hephzibah, Lake Forest Hills, McBean, Monte Sano, National Hills, Rollins, Southside, Willis Foreman, and Windsor Spring Elementary Schools; to the School Nutrition Program; and to the Richmond County Board of Education in Georgia for allowing data to be collected.

Direct observation of meals is often considered the “gold standard” by which dietary assessment tools are validated (1-3) because it is practical (4), economical (3), independent of the subject's memory (5,6), and can provide unbiased information about the subject's actual intake (7). For direct observation of meals, observers typically watch subjects throughout a defined period (eg, school lunch) and take notes on the subjects' eating behaviors regarding items and amounts consumed, traded (received/given away), and/or spilled. Examples of 26 studies in which direct observation has been used include studies to validate dietary recalls (5,6,8-22), food frequency questionnaires(11,23), and food/diet records (11,24) as well as studies to assess dietary intake of individuals (25-31) and to evaluate nutrition education interventions (31).

The use of direct observation as a validation tool is based on the assumption that what is observed is a reliable and valid measure of actual dietary intake (22). Although many studies utilize several people to conduct direct observations, few published studies indicate whether interobserver reliability (IOR) was assessed. According to Simons-Morton and Baranowski (2), who refer to IOR as interobserver agreement (IOA), “IOA compares records of two simultaneous observations on identification of foods or amounts of each food eaten by a subject.” Assessment of IOR is necessary to determine that the information collected does not depend on who conducted the observation because variation here may be inappropriately construed as error in subjects' self-reports of diet(1). Adequate IOR may be defined as at least 85% agreement (23,27). Therefore, assessment of IOR reflects the level of consistency between observations of the same subject(s) by different observers(1,32,33). This article provides a general review of IOR and describes our method for assessing IOR for a series of three studies.

General Review of Assessment of IOR

Assessment of IOR is important during training of new observers prior to data collection, during data collection among new and experienced observers, and during retraining of observers if levels of agreement are inadequate.

During Training. The goal of IOR assessment during training is slightly different than during data collection. While new observers are being trained regarding the observation protocol, they are usually paired with experienced observers to assess IOR and to determine whether the observations of the new observer are consistent with those of the experienced observer(s). After several IOR assessments during training, if new observers fail to meet the level of agreement needed for data collection, further training is needed. Data collection should not begin until adequate agreement is achieved. Of the 26 examples of studies that used direct observation and were cited previously, only three (20,27,31) indicated assessment of IOR during training.

During Data Collection. Once data collection begins, IOR should be assessed among all observers on a regular basis (eg, weekly) to ensure that the level of agreement is maintained throughout the entire data collection period. If at any point during data collection IOR is inadequate, then the IOR assessment in question should be evaluated and discussed. If this occurs consistently, data collection should be suspended so that observers can be retrained and IOR reassessed until adequate agreement can again be achieved. Otherwise, confidence cannot be placed in any of the observations because it cannot be said which, if any, reflects the subject's actual intake. Regarding the 26 studies mentioned previously, two (15,16) specified having only one observer; thus, assessment of IOR was not possible. Another study (29) mentioned assessment of IOR, but it is unclear (a) whether the observations were conducted at school and home as during data collection,(b) how many people conducted the observations and were involved in IOR, and (c) when IOR was done (ie, during training and/or data collection). One study (9) assessed IOR between an observer and an additional observer who *only* conducted IOR observations. Only 10 (5,6,19,20,23,24,26,27,30,31) of the remaining 22 studies

mentioned assessment of IOR during data collection and provided methodologic information and/or results for IOR; three of these studies (5,6,20) were conducted by our group and are discussed further in this article.

As indicated in Figure 1, in each of the remaining seven studies of the 10 studies mentioned previously, the frequency for assessing IOR during data collection, if specified, depended primarily on the study's observation protocol. For example, when observations each covered a 12-hour period, IOR was assessed during either two 20-minute (19) or three 30-minute (23) periods each day. IOR was assessed weekly in one study (31) and on randomly selected days over 2 months in another (27). Assessment of IOR during data collection in these seven studies, if specified, was based on agreement on one variable or a combination of variables, such as nutrient intake per meal (27), portion sizes of food/beverage items served (19), and amount or number of servings of food/ beverage items eaten (19,24,26). In summary, many studies that use direct observation fail to mention assessment of IOR; of those that do, the frequency for assessing IOR and the specific variables assessed vary considerably as shown in Figure 1.

In addition to the general review of IOR just provided, this article describes our method for assessing IOR for items and amounts consumed during school meals for a series of studies (referred to as study A, study B, and study C) regarding the accuracy of fourth-grade children's dietary recalls validated with direct observation of school meals (breakfast, lunch) (5,6,20). School meals are consumed by millions of children in the United States each school day. For example, each school day in 2002, the School Breakfast Program served an average of 8.1 million children (34), and the National School Lunch Program served an average of 28.0 million children (35). Meals eaten in group settings such as school, worksite, and camp cafeterias provide excellent opportunities to validate dietary recalls in several population groups (1,3). In these group settings, food/beverage items are usually served using standardized portions and can be easily identified, which facilitates observation. Although the method for assessing IOR described in this article was used in elementary school cafeterias, it can be adapted to other group settings as well.

METHODS AND RESULTS

Description of Studies A, B, and C

In three consecutive studies, randomly selected fourth-grade children from one school district were observed eating school meals (breakfast, lunch) by research dietitians and interviewed later regarding their dietary intake. The Institutional Review Board approved each study. Child assent and parental consent were both obtained for study participants. Research dietitians observed one to three children simultaneously by standing near tables at which groups of children sat and recorded what they ate on an observation form. The amount eaten for each item was recorded in servings as all (100%), most (75%), half (50%), little bit (25%), taste (10%), none (0%), or the actual number of servings eaten if >1 (eg, 1 + most). Children in general knew they were being observed, but individual children did not know who would be interviewed later. The observations of school meals were used to validate these portions of the children's dietary recalls. Data were collected on children from six elementary schools in study A (20), 11 elementary schools in study B (5), and 10 elementary schools in study C (6). Information regarding interview procedures for each study, quality control for interviews, and results regarding the accuracy of children's dietary recalls is described elsewhere (5,6,20,36).

Frequency of IOR Assessment for Studies A, B, and C

During Training. IOR during training for each study had to be acceptable (ie, at least 85% mean agreement across observers for amounts of food/beverage items eaten across all subjects) prior to initiating data collection. For study A, all three observers were new; during training,

IOR was assessed on a total of 5 days. For study B, there were two new observers and two experienced observers from study A; during training of new observers, IOR was assessed on a total of 10 days between all pairs of new and experienced observers. For study C, there was one new observer and two experienced observers from study B; during training of the new observer, IOR was assessed on a total of 7 days between all pairs of new and experienced observers.

During Data Collection. IOR was assessed twice monthly for study A (20), during 12 of 16 weeks for study B (5), and weekly for study C (6). IOR was usually assessed on Fridays for all three studies so that the other school days of the week could be used for actual data collection (ie, Mondays through Thursdays for observations of school meals and Monday evenings through Friday mornings for dietary recall interviews). For each study, whenever possible, IOR assessment took place equally across all schools at which data were collected.

Assessment of IOR for Studies A, B, and C

When assessing IOR, each participating observer used the same standardized observation form that was used for non-IOR observations. Typically, IOR was assessed on two children per meal. In study A, all three observers assessed IOR simultaneously. However, having three observers in the school cafeteria simultaneously observing the same one or two children in one class at a meal for IOR was cumbersome; the observers often found it difficult to find places to stand where they could adequately observe the selected children without seeing what the other observers were writing on their observation forms. Toward the end of study A, one observer resigned, after which IOR was assessed between pairs of observers, which continued throughout studies B and C. IOR was not assessed on a particular child more than one time. In addition, children observed for IOR were not interviewed to obtain a dietary recall regarding those meals, but they could be observed (and interviewed) on another day for data collection.

On a specific day when IOR was assessed, the observers participating in IOR watched the same children during both breakfast and lunch. Before breakfast began that day, the observers decided which class they would observe. The first child to arrive for breakfast in the agreed upon class was the first child observed for IOR (unless he/she had been observed previously for IOR). Observers participating in IOR used subtle nods or whispers to reach agreement regarding which child was to be observed and made every effort to be as unobtrusive as possible to lessen reactivity (2). One or all of the observers participating in IOR could reposition themselves as necessary to see what the child was eating; however, observers were not to look at what the other observer(s) wrote on her/their observation form(s). Observations were conducted according to a standardized protocol (5,6,20,24). While the first child was being observed at breakfast, the observers identified an additional child to observe from the same class. At most schools, children sat according to the time they arrived at school at breakfast but by class grouping at lunch. The two children on whom IOR was assessed at breakfast on a given day had to be in the same class; this was because children were observed for both breakfast and lunch on an individual day and because only one class could be observed at lunch later that day due to staggered lunch schedules by class. After an observation for IOR was complete, each observer participating in IOR was required to fill out her form completely before comparing her form to other observers' forms; changes could not be made to forms once they were compared.

For study A, for each meal observed for a child on whom IOR was assessed, percentage agreement was calculated for amounts observed eaten within one-quarter serving across all participating observers. To facilitate this process, for studies B and C, an IOR checklist was developed for evaluating the IOR observations. The IOR checklist contained various quality control checks and a space at the bottom to calculate percentage agreement between the observers regarding amounts observed eaten. For each child on whom IOR was assessed, an

IOR checklist was completed by each observer in the pair, as well as other observers for the study and the principal investigator (collectively referred to as “team members”).

On the IOR checklist, each team member noted discrepancies between the observation forms completed by the observers in the pair. Next, team members calculated the “number of items” by tallying items observed by one or both observers in the pair. The amounts eaten recorded by each observer in the pair were compared, and the “number of items that agreed within one-quarter serving” was tallied. As shown in the example in the Table, during a breakfast observation, of the four items observed on the child's tray by one or both observers, the observers agreed within one-quarter serving on three items (cereal, milk, and graham crackers). During the lunch observation, of the seven items observed on the child's tray by one or both observers, the observers agreed within one-quarter serving on six items (hot dog, french fries, green beans, apple, milk, and ketchup). Percentage agreement was calculated by dividing the “number of items that agreed within one-quarter serving” for both meals by the total “number of items.” Clearly, as the number of items varied for each day of assessment, the penalty for lack of agreement within one-quarter serving could be lesser or greater. For example, IOR would be 78% for a day with nine items in which there was agreement for seven items, and IOR would be 83% for a day with 12 items in which there was agreement for 10 items.

After all IOR checklists were completed on a specific day, they were reviewed by all team members. If IOR for a specific child for both school meals was inadequate (ie, percentage agreement was less than 85%), it was discussed during the next weekly team meeting to determine how to correct the problem(s). If IOR was inadequate consistently (ie, for consecutive assessments on more than three children), data collection would be stopped and further retraining would occur.

Results from IOR Assessed during Data Collection for Studies A, B, and C

Study A. IOR was assessed on 9 days on 17 children from five schools during 14 weeks of data collection. Figure 2 indicates mean percentage agreement for each of the 17 children in the order for which IOR was assessed for study A. There was 92% mean agreement (median=91%, minimum=77%, maximum=100%) across three observers for food items in which the amounts observed eaten were within one-quarter serving. (The mean percentage agreement has increased from 89% since the results of study A were published (20) because several errors were reconciled while writing this article. Furthermore, one IOR assessment was deleted from the original 18 IOR assessments because the child had been observed for IOR previously.) On each of the final 2 days on which IOR was assessed, there were two observers instead of three. For study A, the IOR assessment for subject 8 (Figure 2) had the lowest mean percentage agreement at 77%; the three observers agreed within one-quarter serving on 10 of 13 items. The three items with discrepancies occurred during the lunch observation. For all three items, two of the three observers were in agreement. In addition, one observer did not record one of these items. All subsequent IOR assessments for study A had a greater mean percentage agreement that was within the acceptable limits (at least 85%).

Study B. IOR was assessed on 12 days on 22 children from nine schools during 16 weeks of data collection. Figure 3 indicates mean percentage agreement for each of the 22 children in the order for which IOR was assessed for study B. There was 90% mean agreement (median=91%, minimum=73%, maximum=100%) across pairs of observers for food items in which the amounts observed eaten were within one-quarter serving (5). For study B, the IOR assessment for subject 12 (Figure 3) had the lowest mean percentage agreement at 73%; the two observers agreed within one-quarter serving on eight of 11 items. Amounts for three of the four items from the breakfast observation were in agreement; one observer did not record a trade, which altered the amount eaten recorded on her observation form for one item. During the lunch observation, a taco salad was served to the child; one observer listed it as one item,

whereas the other marked the salad separate from the meat and taco chips. Therefore, the number of items recorded varied by one, and the amounts eaten were different for two items. All subsequent IOR assessments for study B had a greater mean percentage agreement that was within the acceptable limits (ie, at least 85%), except for the final one, which was 82%.

Study C. IOR was assessed on 6 days on 10 children from six schools during 6 weeks of data collection. Figure 4 indicates mean percentage agreement for each of the 10 children in the order for which IOR was assessed for study C. There was 93% mean agreement (median =92%, minimum =80%, maximum =100%) across pairs of observers for food items in which the amounts observed eaten were within one-quarter serving (6). For study C, the IOR assessment for subject 6 (Figure 4) had the lowest mean percentage agreement at 80%; the two observers agreed within one-quarter serving on eight of 10 items. Amounts for three of the four items from the breakfast observation were in agreement, and amounts for five of the six items from the lunch observation were in agreement. In each discrepancy, the amount observed eaten differed by half a serving. All subsequent IOR assessments for study C had a greater mean percentage agreement that was within the acceptable limits (at least 85%).

Summary. Mean percentage agreement for IOR during data collection for studies A, B, and C ranged from 90% to 93%. IOR during data collection was never unacceptable for more than three consecutive children in any of the studies. Although there were a few subjects for which IOR was unacceptable, overall, mean IOR was at least 85% and, therefore, acceptable (23, 27). Retraining of observers was not necessary during data collection for any of the three studies.

DISCUSSION

An important strength of our method for assessing IOR used in these studies is that IOR was assessed during training as well as regularly throughout data collection. Thus, observers were encouraged to remain alert and follow the observation protocol to maintain adequate IOR levels throughout the entire data collection period instead of only during training prior to data collection. Another strength is that IOR was assessed for foods at the item and amount level, as opposed to the nutrient level. This was done because observers watch children eating *foods*, not nutrients. For example, if one observer records that a child drank apple juice, but the other observer records that the same child drank orange juice instead, IOR assessed at the nutrient level would identify this as an error for some nutrients (eg, vitamin C and potassium) but not for other nutrients (eg, kilocalories, protein, fat, and carbohydrate). Thus, depending on the nutrient profile for each of the two food items, and the nutrient(s) of interest for a specific research study, an IOR error might, or might not, be evident if IOR is assessed at the nutrient level.

A limitation of our method for assessing IOR is that the observations assessed for IOR were on days that were not used to collect data. Therefore, each observer participating in IOR knew in advance that her observations were being assessed for IOR. To avoid advance knowledge of IOR assessment, more than one observer would have to observe every child for each meal and day of data collection so that some of these days could be randomly selected for IOR assessment. This would severely decrease the number of children that could be observed on an individual day for data collection and thereby inflate to unreasonable levels the budget for personnel to conduct observations. Assessing IOR on a day each week or month that would not impact data collection made optimal use of every day available because more children could be observed on data collection days, and nondata collection days could be used for IOR.

CONCLUSIONS

- The method for assessing IOR described can be applied to nutrition studies that use direct observation either to validate dietary assessment tools or to acquire information about subjects' dietary intake. Assessment of IOR is necessary when multiple people conduct observations for a research study to ensure that the quality of the information obtained does not depend on who conducted the observation.
- Many publications and oral presentations of studies that use observations fail to mention IOR. Researchers who have multiple people conducting observations for nutrition studies need to assess IOR and to include results for IOR in published reports and/or oral presentations.
- Assessment of IOR is important during training of observers, throughout the entire data collection period, and for retraining during data collection as needed. Adequate agreement between observers is essential to the reasoning behind using observation as a validation tool.
- Readers are encouraged to question the results of studies that have multiple people conducting observations but fail to mention assessment of IOR and/or fail to include the results of IOR assessment.

References

1. Frank GC. Taking a bite out of eating behavior: Food records and food recalls of children. *J Sch Health* 1991;61:198–200. [PubMed: 1943041]
2. Simons-Morton BG, Baranowski T. Observation in assessment of children's dietary practices. *J Sch Health* 1991;61:204–207. [PubMed: 1943043]
3. Mertz W. Food intake measurements: Is there a “gold standard”? *J Am Diet Assoc* 1992;92:1463–1465. [PubMed: 1452958]
4. Block G. A review of validations of dietary assessment methods. *Am J Epidemiol* 1982;115:492–505. [PubMed: 7041631]
5. Baxter SD, Thompson WO, Smith AF, Litaker MS, Yin Z, Frye FHA, Guinn CH, Baglio ML, Shaffer NM. Reverse versus forward order reporting and the accuracy of fourth-graders' recalls of school breakfast and school lunch. *Prev Med* 2003;36:601–614. [PubMed: 12689806]
6. Baxter SD, Thompson WO, Litaker MS, Guinn CH, Frye FHA, Baglio ML, Shaffer NM. Accuracy of fourth-graders' dietary recalls of school breakfast and school lunch validated with observations: In-person versus telephone interviews. *J Nutr Educ Behav* 2003;35:124–134. [PubMed: 12773283]
7. Gersovitz M, Madden JP, Smiciklas-Wright H. Validity of the 24-hr. dietary recall and seven-day record for group comparisons. *J Am Diet Assoc* 1978;73:48–55. [PubMed: 659761]
8. Baranowski T, Islam N, Baranowski J, Cullen KW, Myres D, Marsh T, de Moor C. The Food Intake Recording Software System is valid among fourth-grade children. *J Am Diet Assoc* 2002;102:380–385. [PubMed: 11902371]
9. Reynolds LA, Johnson SB, Silverstein J. Assessing daily diabetes management by 24-hour recall interview: The validity of children's reports. *J Pediatr Psychol* 1990;15:493–509. [PubMed: 2258797]
10. Madden JP, Goodman SJ, Guthrie HA. Validity of the 24-hr. recall. *J Am Diet Assoc* 1976;68:143–147. [PubMed: 1245715]
11. Crawford PB, Obarzanek E, Morrison J, Sabry ZI. Comparative advantage of 3-day food records over 24-hour recall and 5-day food frequency validated by observation of 9- and 10-year-old girls. *J Am Diet Assoc* 1994;94:626–630. [PubMed: 8195550]
12. Dubois S, Boivin J-F. Accuracy of telephone dietary recalls in elderly subjects. *J Am Diet Assoc* 1990;90:1680–1687. [PubMed: 2246447]
13. Lytle LA, Murray DM, Perry CL, Eldridge AL. Validating fourth-grade students' self-report of dietary intake: Results from the 5-A-Day PowerPlus program. *J Am Diet Assoc* 1998;98:570–572. [PubMed: 9597031]

14. Baxter SD, Thompson WO, Davis HC, Johnson MH. Impact of gender, ethnicity, meal component, and time interval between eating and reporting on accuracy of fourth-graders' self-reports of school lunch. *J Am Diet Assoc* 1997;97:1293–1298. [PubMed: 9366868]
15. Baxter SD, Thompson WO, Davis HC. Prompting methods affect the accuracy of children's school lunch recalls. *J Am Diet Assoc* 2000;100:911–918. [PubMed: 10955049]
16. Domel SB, Thompson WO, Baranowski T, Smith AF. How children remember what they have eaten. *J Am Diet Assoc* 1994;94:1267–1272. [PubMed: 7963170]
17. Stunkard AJ, Waxman M. Accuracy of self-reports of food intake. *J Am Diet Assoc* 1981;79:547–551. [PubMed: 7026651]
18. Lytle LA, Nichaman MZ, Obarzanek E, Glovsky E, Montgomery D, Nicklas T, Zive M, Feldman H. Validation of 24-hour recalls assisted by food records in third-grade children. *J Am Diet Assoc* 1993;93:1431–1436. [PubMed: 8245378]
19. Baranowski T, Sprague D, Baranowski JH, Harrison JA. Accuracy of maternal dietary recall for preschool children. *J Am Diet Assoc* 1991;91:669–674. [PubMed: 2040780]
20. Baxter SD, Thompson WO, Litaker MS, Frye FHA, Guinn CH. Low accuracy and low consistency of fourth-graders' school breakfast and school lunch recalls. *J Am Diet Assoc* 2002;102:386–395. [PubMed: 11905461]
21. Basch CE, Shea S, Arliss R, Contento IR, Rips J, Gutin B, Irigoyen M, Zybert P. Validation of mothers' reports of dietary intake by four to seven year-old children. *Am J Public Health* 1990;80:1314–1317. [PubMed: 2240296]
22. Carter RL, Sharbaugh CO, Stapell CA. Reliability and validity of the 24-hour recall. *J Am Diet Assoc* 1981;79:542–547. [PubMed: 7288060]
23. Baranowski T, Dworkin R, Henske JC, Clearman DR, Dunn JK, Nader PR, Hooks PC. The accuracy of children's self-reports of diet: Family Health Project. *J Am Diet Assoc* 1986;86:1381–1385. [PubMed: 3760429]
24. Domel SB, Baranowski T, Leonard SB, Davis H, Riley P, Baranowski J. Accuracy of fourth- and fifth-grade students' food records compared with school lunch observations. *Am J Clin Nutr* 1994;59:S218–S220.
25. Davidson FR, Hayek LA, Altschul AM. Towards accurate assessment of children's food consumption. *Ecology Food Nutr* 1986;18:309–317.
26. Hanson SL, Pichert JW. Perceived stress and diabetes control in adolescents. *Health Psychol* 1986;5:439–452. [PubMed: 3757992]
27. Simons-Morton BG, Forthofer R, Huang IW, Baranowski T, Reed DB, Fleishman R. Reliability of direct observation of schoolchildren's consumption of bag lunches. *J Am Diet Assoc* 1992;92:219–221. [PubMed: 1737906]
28. Lorenz RA, Christensen NK, Pichert JW. Diet-related knowledge, skill, and adherence among children with insulin-dependent diabetes mellitus. *Pediatrics* 1985;75:872–876. [PubMed: 3991274]
29. Zive MM, Berry CC, Sallis JF, Frank GC, Nader PR. Tracking dietary intake in white and Mexican-American children from age 4 to 12 years. *J Am Diet Assoc* 2002;102:683–689. [PubMed: 12008994]
30. Simmons SF, Reuben D. Nutritional intake monitoring for nursing home residents: A comparison of staff documentation, direct observation, and photography methods. *J Am Geriatr Soc* 2000;48:209–213. [PubMed: 10682952]
31. Coates TJ, Jeffery RW, Slinkard LA. Heart healthy eating and exercise: Introducing and maintaining changes in health behaviors. *Am J Public Health* 1981;71:15–23. [PubMed: 7258426]
32. Baranowski T, Simons-Morton BG. Dietary and physical activity assessment in school-aged children: Measurement issues. *J Sch Health* 1991;61:195–197. [PubMed: 1943040]
33. Kirk, J.; Miller, ML. *Qualitative Research Methods*. Vol.1. SAGE Publications, Inc; Beverly Hills, CA: 1986. Reliability and Validity in Qualitative Research.
34. US Department of Agriculture, Food and Nutrition Service School Breakfast Program Annual Summary <http://www.fns.usda.gov/pd/sbsummar.htm> Accessed December 17, 2003
35. US Department of Agriculture, Food and Nutrition Service National School Lunch Annual Summary <http://www.fns.usda.gov/pd/slsummar.htm> Accessed December 17, 2003

36. ShafferNM BaxterSD ThompsonWO BaglioML GuinnCHFryeFHA Quality control for interviews to obtain dietary recalls from children for research studies. *J Am Diet Assoc* In press

Figure 1.
Assessment of interobserver reliability (IOR) during data collection for seven studies.

Figure 2. Interobserver reliability (IOR) results in chronological order for study A. (IOR was assessed twice monthly on 9 days during 14 weeks of data collection. Two subjects were observed on each day of IOR assessment except for day 9 on which Subject 17 was the only subject observed for IOR.)

Figure 3.

Interobserver reliability (IOR) results in chronological order for study B. (IOR was assessed weekly on 12 days during 16 weeks of data collection. Two subjects were observed on each day of IOR assessment except for day 2 on which three subjects [Subjects 3, 4, and 5] were observed for IOR and days 5, 9, and 10 on each of which only one subject [Subjects 10, 17, and 18, respectively] was observed for IOR.)

Figure 4. Interobserver reliability (IOR) results in chronological order for study C. (IOR was assessed weekly on 6 days during 6 weeks of data collection. Two subjects were observed on each day of IOR assessment except for days 2 and 5 on which only one subject [Subjects 3 and 8, respectively] was observed for IOR.)

Table
Example of calculation of interobserver reliability for a subject

Item description	Amount Observed Eaten ^a	
	Observer 1	Observer 2
Breakfast		
Cereal, sugar-coated flakes	All	All
Orange juice	All	Half
Milk, 1%	All	All
Graham crackers, honey	Half	Little bit
Lunch		
Hot dog on bun	All	All
French fries	Most	Half
Green beans	None	None
Apple wedges	Little bit	Little bit
Milk, chocolate	All	Most
Ketchup	All	All
Mustard	All	(Not observed)

Calculation of interobserver reliability based on items listed above that were observed at breakfast and lunch for this subject:

Number of items (ie, tally of items observed by one or both observers in the pair)=**11**.

Number of items that agreed within one-quarter serving=**9**.

Percent agreement= $(9 \div 11) \times 100 = \mathbf{82\%}$.

^a Amount eaten was recorded in servings as all=100%, most=75%, half=50%, little bit=25%, taste=10%, none=0%, or the actual number if >1 (eg, 1 + most).