New technology in dietary assessment: a review of digital methods in improving food record accuracy

Phyllis J. Stumbo
Institute for Clinical and Translational Sciences, University of Iowa, 1700 First Avenue Suite 17, Iowa City, IA 52242, USA

Methods for conducting dietary assessment in the United States date back to the early twentieth century. Methods of assessment encompassed dietary records, written and spoken dietary recalls, FFQ using pencil and paper and more recently computer and internet applications. Emerging innovations involve camera and mobile telephone technology to capture food and meal images. This paper describes six projects sponsored by the United States National Institutes of Health that use digital methods to improve food records and two mobile phone applications using crowdsourcing. The techniques under development show promise for improving accuracy of food records.

Dietary assessment: Dietary record: Mobile telephone: Computer-assisted image processing: Epidemiological methods

The first written report of dietary assessment on the western side of the Atlantic appeared in the 1930s. Medlin and Skinner(1) reviewed published reports of dietary assessment between about 1930 and 1985 and found early discussions of daily food intake being recorded by clinic patients and dietitians in both the US and Europe, primarily Great Britain. In the US where kitchen scales were not common portions recorded were estimated, but in England there was a tendency to weigh the food. Soon nutritionists discovered that a different answer might result depending on how the dietary intake data were collected. Even with good records, another challenge was determining the composition of foods recorded. This led to short methods to manually calculate composition of dietary intake records(2). This manual calculation process was replaced by computerised systems in the 1970s and 1980s as computers became more accessible.

Numerous attempts were made to determine the accuracy of food intakes recorded by research participants and how to determine how many days of information were required to accurately reflect usual intake. Basiotis et al.(3) with the US Department of Agriculture created an interesting dataset consisting of intake records kept daily for an entire year by twenty-nine study volunteers in Beltsville, MD. By analysing the composition of these records the authors found that a single day’s intake, or even records kept for a week or longer, were not representative of ‘usual intake’, and it was the usual intake that was most useful for dietary assessment. Intakes of energy and protein were relatively stable from day to day. Thus, an estimate of 3 or 4 d intake was considered to accurately represent the intake for a year for a group of people, but according to their work the individual needed to keep records for a month to accurately reflect usual intake of protein or energy. And if this was not daunting enough, nutrients that are concentrated in a few foods were even more difficult to assess, e.g. records kept for an entire year still did not accurately reflect an individual’s usual intake of vitamin A.

Fortunately, since that time mathematical models have been developed to estimate usual intake using fewer days of intake(4,5).

For most research goals, long-term (usual) intake is of primary interest but clients are seldom willing to record food intake for long periods of time. Further complicating the problem, the process of calculating composition of the food records is expensive and time-consuming making it
even less feasible to determine any one persons ‘usual’ intake by relying on recalls or records of an individual’s daily intake. To solve these problems, short questionnaires called FFQ have been developed to summarise intake over extended periods as well as computerised procedures for determining composition of the intakes without manually coding each food. The other motivator to develop the FFQ was the desire to collect dietary information in epidemiological studies with large study samples in locations remote from the study centre. The FFQ could be mailed and was low cost. Many FFQ have been developed since the first ones introduced in the 1960s. For purposes of this paper, the FFQ developed in 1986 by a team of nutritionists at the US Department of Health and Human Services will be highlighted. Using data from the National Health and Nutrition Examination Survey these nutritionists selected about 147 food items or food groups that represented 90% of the total intake of eight nutrient-dense foods among all persons surveyed, including vitamins A and C that are concentrated in a few foods. From these data an FFQ was developed. This FFQ and others have been used in numerous nutrition studies. However, the problem of accurately assessing long-term intake was not adequately solved.

Many researchers doubted the ability of human subjects to accurately estimate their average or usual intake for an entire year. However, a major problem in validating intake questionnaires was a lack of a gold standard to judge them. The best tools available were the dietary record or the 24-h dietary recall. A major limitation of these methods was the correlated error between the methods which could result in spuriously high correlations, thus the need to consider unbiased biomarkers was introduced.

A new study was mounted to evaluate whether FFQ in use represented ‘usual’ intake as measured by food recalls. Recognising that food recalls give variable results they collected data for 1 d in each of four seasons, hoping this amount of data would demonstrate the utility of the FFQ.

What it did was confirm their concern over inaccuracy of the FFQ. Energy correlated at less than 0.5 with energy requirements of this weight-stable population, and calculating an ‘energy adjusted’ value brought correlations to only slightly greater than 0.5. This study confirmed suspicions that the FFQ as it was constructed lacked the precision needed for research about health consequences from dietary factors. The confirming data came later, when a study dubbed the ‘OPEN’ study used doubly labelled water as a biomarker for energy intake and urinary N for protein intake found that subjects underreported energy and protein intake by approximately one third on the FFQ.

Clearly researchers interested in studying the effect of lifestyle factors on disease required more valid ways to study dietary intake. Whether records or recalls were to be the primary study method, or used to support development of more efficient FFQ, improved methodology was needed. Thus a call for proposals for improved methods for capturing diet and physical activity data was issued. The following describes projects resulting from that request that are under development using advanced technology to capture food intake data. The confirmatory data came later, when a study dubbed the ‘OPEN’ study used doubly labelled water as a biomarker for energy intake and urinary N for protein intake found that subjects underreported energy and protein intake by approximately one third on the FFQ.

As early as the 1960s, the most common use of pictures for diet assessment require observers to make subjective estimates of the amounts consumed compared with quantities shown in the image. For example, Fig. 1 shows two-dimensional food models that display several serving sizes of common foods and the user chooses the picture that best matches the amount they recall having eaten.

Cameras have also been used to record meals before and after eating as a visual food record. The advent of disposable cameras gave a boost to their use in this latter method of documenting food intake. Several academic projects demonstrated usefulness of this technique, but carrying a camera at all times, the inconvenience of mailing the camera to researchers, and the lack of immediate feedback made the process burdensome and undesirable. The eventual appearance of mobile telephones with cameras and wireless transmission greatly increased attractiveness of images as a food record research tool. However, a picture by itself usually requires additional information, and unless the picture-taking is standardised it is difficult to identify a food and estimate its portion size. The National Institutes of Health dietary assessment projects have aimed to correct some of these problems.

Some of the technical details involved in using a camera image of a meal to accurately measure the food volume and type are described here. Of the six applications in Table 1, two rely entirely on self-report to identify foods and portions consumed, i.e. a respondent identifies the food by name from a database and the serving size by comparison with graduated photographs (Subar and Baranowski projects). The other four projects use images of the food to be consumed and to varying degrees utilise automatic procedures for identifying and quantifying foods in the

Table 1. US National Institutes of Health Sponsored Technology-Assisted dietary assessment projects

<table>
<thead>
<tr>
<th>Investigator</th>
<th>Project title</th>
<th>Project nickname</th>
</tr>
</thead>
<tbody>
<tr>
<td>T. Baranowski</td>
<td>Food Intake Recording Software System</td>
<td>FIRSTST</td>
</tr>
<tr>
<td>C. Boushey</td>
<td>Improving Dietary Assessment Methods Using the Cell Phone and Digital Imaging (Technology Assisted Dietary Assessment)</td>
<td>TADA</td>
</tr>
<tr>
<td>A. Kristal</td>
<td>Integrated sensor technology for real-time recording of food intake (Dietary Data Recorder System)</td>
<td>DRRS</td>
</tr>
<tr>
<td>A.F. Subar</td>
<td>Automated Self-Assisted 24-h Dietary Recall</td>
<td>ASA24</td>
</tr>
<tr>
<td>M. Sun</td>
<td>A Unified Sensor System for Ubiquitous Assessment of Diet and Physical Activity</td>
<td>eButton</td>
</tr>
<tr>
<td>R. Weiss</td>
<td>Mobile Food Intake Visualisation and Voice Recogniser (FIVR)</td>
<td>FIVR</td>
</tr>
</tbody>
</table>
image or employing trained analysts to estimate amount consumed from images of the foods taken before and after eating.

One of the big hurdles in this process is automatically determining serving size (or volume) from images. To measure volume from a flat image more than one pose from a scene is usually required to reveal the shape of the food. This is accomplished by moving the camera over or around the food before eating and clicking the camera three or more times during the arc. This sounds simple, but actually requires a steady hand to press the shutter without shaking the camera thus blurring the image. An alternative method is to use the video function and capture continuous images throughout an arc with the camera automatically opening the shutter as the food is scanned.

While recognising the advantage of using more than one picture to measure volume or serving size, two of the projects (eButton and Technology Assisted Dietary Assessment (TADA)) have developed algorithms to use a single image in calculating volume believing that capturing a single image will be easier for consumers. The mathematical procedures involved in calculating volume of the foods within the image vary between these two projects. eButton is a device worn on the chest (like a pin) that contains a miniature camera, accelerometer, global position system and other sensors and captures data and information of health activities, eliminating the need for daily self-reporting\(^{(15)}\). When analysing video capture of a meal eButton creates a virtual scene where each component of the meal captured by the video camera is enclosed with a triangle mesh\(^{(16)}\). Figure 2 illustrates the use of a triangle-based polygon representation (known as ‘triangle mesh’) to estimate food volume. There are technical details beyond the space allotted here for the discussion. For example, the camera must be calibrated, and in practice it may be necessary to calibrate every device used to document intake with this technique. However, the concept is easy to illustrate. Using one image of a food, a virtual wire frame is selected that resembles the shape and will surround the food allowing the analyst to manipulate the wire frame visually encapsulating the food allowing the volume of be estimated from the size of the triangles forming the enclosure. Figure 2 illustrates how a wire frame can be manipulated on the virtual image of a plate of spaghetti to engulf the food providing measurements needed to estimate volume. In tests to date this works very well with a cube such as cake or cornbread (error 1–10 %), but is more difficult with irregularly shaped foods such as chicken breast (error 21–27 %) or small items such as an onion slice (error 12–17 %)\(^{(16)}\). The manual operation of the mesh eliminates much of the high-level programming required by the more automated applications, but the
increased manual effort may limit its use for large population studies.

The TADA project also uses a single image to estimate serving size, however, eButton has the advantage of selecting the best image from its video capture. The key to calculating accurate measurements from an image lies in use of a standard or fiducial marker within the image. TADA uses a credit card-sized fiducial marker in the form of a \(5 \times \frac{7}{12} \text{cm}^2\) checkerboard square. The marker provides scale for measurements of other objects in the image and a flat plane to estimate the camera orientation. eButton uses the plate to judge actual size of the food, so a plate of known size is required in the image which may further limit its use in large-scale studies.

Several steps are involved in deriving volume measurements. The first is to segment the image into the individual food items. TADA attempts this task automatically using connected ‘component analysis’, ‘active contours’ and ‘normalised cuts’, engineering terms for procedures used in segmenting the image into individual foods for subsequent identification and size estimation. Details of segmentation have been previously reported\(^{18–21}\). Component analysis begins with separating the plate from table cloth and proceeds on to outlining the food components present on the plate. This is most successful when food items do not overlap. A segmented food scene is depicted in Fig. 3.

When the location and identification of each food are established, the volume can then be estimated. The volume estimation process used by TADA involves classifying each segment into a geometric class (sphere, cube and mound), deriving measurements from the image (radius for sphere, length and width for cube, etc.) and applying the appropriate formula to calculate volume. The calculations could be conducted on the handheld device, however, since the calculations would drain battery power the most common procedure will be to send the image to the server for processing. Results of the accuracy of food volume estimations using the TADA device have been reported\(^{21,22}\). Even young TADA users have found the single image capture to be quick and easy to use.

Food Intake Visualisation and Voice Recogniser uses unique visual techniques to calculate volume of food images captured by camera embedded in a mobile telephone by selecting three images from a video recording of the food plate made by moving the camera device in an arc over the meal. Food volume is determined by extracting a three-dimensional point cloud (another engineering term) of all food on the plate (see Fig. 4). Volume estimation is a two-step process, first Delaunay triangulation is performed

---

**Fig. 2.** (colour online) Spaghetti volume estimated using wire mesh: (a) spaghetti on plate, (b) wire mesh selected to match spaghetti shape and (c) wire mesh adjusted to size of spaghetti mound. Image courtesy: Mingui Sun, University of Pittsburg, Pennsylvania, USA, 2012.

**Fig. 3.** (colour online) Segmentation of an image captured by an adolescent during a controlled feeding study. (a) Ground truth segmentation and (b) results of automated segmentation. Images courtesy of VIPER Laboratory, Purdue University, West Lafayette, IN, USA, 2012.
to fit the surface of the food and the total area of each triangle is then determined and used in calculating volume. A food recognition feature is applied to segment the image before volume determination can be performed, thus separating the meal scene into individual foods. Food identification begins with a voice recording by the user naming each food. The food names are then submitted to a software classifier that compares the spoken label plus the colour and texture features in the image to other images from a library of foods with similar names. When a food is matched to one of those in the library, it is segmented based on similarity of colour and texture of each section or patch of the segmented food image in the meal scene. Technical details of the process are described by Puri et al. This application also applies a colour checker-board fiducial marker to guide interpretation of the image.

The Dietary Data Recorder System adds a laser beam to the telephone with embedded camera to avoid the necessity of carrying a fiducial marker. The device is constructed from a standard mobile telephone with a laser generator attached using a custom-built housing depicted in Fig. 5(a). The laser beam projects a light grid onto the food scene intermittently as a guide to size of the scene, and unobstructed images between the intermittent laser grid projections are used by nutritionists who later identify the food and calculate the composition. The images are captured using the video function of the camera while slowly moving the camera around a food (see Fig. 5(b) and (c)). Since the motion between two adjacent frames is small, non-grid images can assist in the analysis of the grid images. The mobile application was developed in JAVA on the Android operating system using the Google Nexus One smartphone. Nexus One is also equipped with a three-axis accelerometer and a digital compass, whose data can be used in the reconstruction process.

These four National Institutes of Health sponsored applications use vision techniques in novel ways to document food intake. They hold promise for more accurate
and objective measurement for food records. The remaining two projects (automated self-administered 24-h dietary recall (ASA24) and Food Intake Recording Software System (FIRSSST)) make use of more traditional methods of data capture while tapping into the internet for improved access and seamless analysis. The ASA24 is a web-based 24-h dietary recall system available from anywhere in the world where there is an internet connection. ASA24 is currently being considered for measuring intakes from other parts of the world is feasible. An adaptation of ASA24 designed for children. ASA24 uses food images to aid in reporting type and amount of foods eaten. The application uses a three-panel food entry screen, two panels on the left and centre guide users to search for foods from a database and a serving-size screen on the right is used to record amount eaten. A penguin avatar gives verbal instructions to augment the written instructions throughout data entry.

The food databases for all these applications currently represent foods as eaten in the US, but adding databases from other parts of the world is feasible. An adaptation of ASA24 is currently being considered for measuring intakes in Canada and Great Britain.

Other techniques for using mobile telephone with a camera to record intake are available. Meal Snap is an application for the iPhone that offers almost immediate feedback about the energy value of food images sent to the provider. Food images are evaluated for energy content by crowdsourcing. At this time the only information provided is an estimated range of energy values for the food imaged (see Fig. 6 for an example). In an informal test of the tool two food images were submitted to the Meal Snap application, a partially eaten egg roll and a plated meal consisting of chicken, vegetables and rice. The half egg roll energy value from Meal Snap of 230–368 kJ (58–88 kcal) agreed with the US Department of Agriculture Food and Nutrient Database for Dietary Studies (FNDDS) value for whole egg roll of 674 kJ (161 kcal); however, the Meal Snap value of 766–1151 kJ (183–275 kcal) for the plated meal does not agree with the database estimate of 302 kJ (720 kcal) probably because the plate image was taken from a distance to capture the full 3/4 cup meal, but appeared to the provider as a much smaller portion, emphasising the importance of using a fiducial marker to give the image scale. Further, in the case of the egg roll, the amount reported from the Meal Snap was the amount of energy in the remaining food. There was no information provided about the amount actually eaten. In the case of the chicken meal, the energy estimate was for the entire amount served. Thus, these tools are interesting; however, the software is not designed to estimate amounts consumed, which is of value to researchers and healthcare practitioners. The applications designed to capture images for full dietary assessment as envisioned by the National Institutes of Health projects have been found to be very acceptable to both adolescents and adults.

An application similar to Meal Snap called Platemate was developed by Noronha et al. at Harvard University in Boston, MA. Both applications utilise crowdsourcing to provide nutritive information about food images provided by users. Crowdsourcing refers to a technique where a question is posed to a group of people who have agreed to provide answers. The people are selected by the developer and their responses are usually tracked and may be rated by the developer. The supposition is that crowdsourcing uses untrained personnel recruited from the general public, but some applicants may possess considerable skill in the topic. Noronha et al. do not explain how their applicants were selected, but they did find that crowdsourcing was nearly as accurate as professional nutritionists in assessing energy and macronutrient content of food from a digital image.

In summary, innovation in diet and lifestyle assessment comes from many sources. The field has come full-circle in the past century, from ‘homely’ methods to record daily food records in the early 1900s to automated and web-enabled systems to do the same today. The field has gone from hand calculation of nutrient intake in the first half of the last century to computer-assisted calculations in the second half. FFQ were developed to simplify data collection and make accumulation of vast amounts of dietary intake data possible. FFQ have their place in dietary research as practical and affordable, but do not provide details about dietary intakes. FFQ, however, may be combined with other short-term instruments such as records or recalls to capture foods that are rarely consumed. It may be possible to draw on the strength of both long- and short-term instruments to improve overall dietary assessment.

The crowdsourcing applications bring to mind the old debate between clinical and actuarial judgment. The clinician integrates objective data and practical experience as they conduct their work. Research more often than not improves the clinician’s success by generating actuarial data that supplement and improves the clinician’s judgment. Crowdsourcing relies on clinical judgment. Technology has added objective data upon which clinical judgment can be based. The greater precision of newer technologies holds promise to give dietary assessment another giant leap forward.

Acknowledgements

The author sincerely thanks Dr Amy Subar, National Cancer Institute, US National Institutes of Health, Bethesda, MD, USA and Dr Carol Boushey, Cancer Center of Hawaii, Honolulu, Hawaii, USA for their thoughtful comments on this manuscript. Also many thanks to Dr Ed Delp...
References

24. Janet Cade – Centre for Epidemiology and Biostatistics, School of Food Science and Nutrition, University of Leeds, Leeds, Great Britain.