The validity of dietary assessment in general practice
Paul Little, Jane Barnett, Barrie Margetts, Ann-Louise Kinmonth, John Gabbay, Rachel Thompson, Daniel Warm, Hilary Warwick, Steven Wooton

Abstract
Objective—To validate a range of dietary assessment instruments in general practice.

Methods—Using a randomised block design, brief assessment instruments and more complex conventional dietary assessment tools were compared with an accepted “relative” standard—a seven day weighed dietary record. The standard was checked using biomarkers, and by performing test-retest reliability in additional subjects (n=29).

Outcomes—Agreement with weighed record. Percentage agreement with weighed record, rank correlation from scatter plot, rank correlation from Bland-Altman plot. Reliability of the weighed record.

Setting—Practice nurse treatment room in a single suburban general practice.

Subjects—Patients with risk factors for cardiovascular disease (n=61) or age/sex stratified general population group (n=50).

Results—Brief self completion dietary assessment tools based on food groups eaten during a week show reasonable agreement with the relative standard. For % energy from fat and saturated fat, non-starch polysaccharide, grams of fruit and vegetables and starchy foods consumed the range of agreement with the standard was: median % difference −6% to 12%, rank correlation 0.5 to 0.6. This agreement is of a similar order to the reliability of the weighed record, as good as or better than test standard agreement for more time consuming instruments, and compares favourably with research instruments validated in other settings. Under-reporting of energy intake was common (40%) and more likely if subjects were obese (body mass index (BMI) ≥30 60% under-reported; BMI <30 29%, p<0.001).

Conclusion—Under-reporting of absolute energy intake is common, particularly among obese patients. Simple self assessment tools based on food groups, designed for practice nurse dietary assessment, show acceptable agreement with a standard, and suggest such tools are sufficiently accurate for clinical work, research, and possibly population dietary monitoring.

Since the 1990 contract for general practice, practice nurse numbers and their role have increased dramatically. In addition to conventional nursing roles, an important development has been their role in giving lifestyle advice to patients at risk of cardiovascular disease. Although there is good evidence that modification of diet and physical activity are important in many of the conditions managed in primary care—particularly for established ischaemic heart disease, hypertension, and hypercholesterolaemia—it is not clear whether this evidence can be extrapolated to general practice settings. Of particular concern is whether health professionals—and in particular practice nurses—have the knowledge and skills to be able to effectively assess and counsel such patients about their diet.

Specifically, what is the evidence for the effectiveness of practice nurses in providing dietary counselling? The British Family heart and OXCHECK studies demonstrated moderate cost effectiveness comparable to many medical interventions. However, the cost effectiveness was critically dependent on practice nurse time, a major part of which was spent performing dietary assessment. If time spent on dietary assessment is a significant determinant of cost effectiveness, then it is vital to demonstrate the utility and validity of different ways of assessing diet that have very different time requirements. Furthermore, if dietary assessment is to be used in everyday clinical work in primary care, or for research, it should be subjected to validation as are other clinical measurements, for example, blood pressure, weight or peak flow: inaccurate dietary assessment will lead to the wrong patients being targeted and result in wasteful and inappropriate counselling.

What is the evidence for the validity of dietary assessment tools in primary care? There are many brief unvalidated instruments, and equally many validated instruments for research purposes, but only one brief instrument validated in the UK (in an occupational setting). No assessment instruments administered by health professionals have been validated among patients from general practice: the subjects and setting of dietary assessment are likely to affect perceptions, behaviour, and responses. This leaves significant doubts about the validity and efficiency of practice nurse led dietary assessment. The assessment of the relative validity of dietary assessment tools in primary care is thus needed for a range of potential uses—clinical work, research, and possibly population monitoring as most of the population attend general practice each year.
We report the comparative validation characteristics of dietary assessment tools for clinical and research purposes, when administered by a practice nurse to patients in a primary care setting.

Methods

SETTING
Practice nurse treatment room in a single suburban practice in Southampton.

ASSESSMENTS
Assessments were all performed by a practice nurse in the treatment room, using set written protocols to maximise the generalisability of the assessments. The practice nurse was trained in the supervision of the weighed record by a nutritionist and a dietitian (SW, HW).

SAMPLE SIZE
We estimated from previous studies that to estimate true energy intakes to within 4% of a reference method requires about 100 subjects, or to assess intake within 6% 50 subjects. Assuming a test method could predict people consuming more than 33% energy from total fat with a sensitivity of 70%, to estimate the sensitivity with confidence intervals of ±10% requires 100 subjects and with confidence intervals of ±15% 50 subjects (CIA BMJ Publications). Thus, a minimum of 100 subjects were required for the main study.

PATIENTS
We aimed to recruit a minimum of 100 patients by two mechanisms:

1. **High risk group (n=61)**
   Patients at higher risk of cardiovascular disease with no previous counselling by the practice nurse about lifestyle. These are the patients that GPs and practice nurses are most likely to prioritise for dietary assessment and counselling. This group was recruited opportunistically when attending surgery for other reasons. Criteria for inclusion were: one risk factor for cardiovascular disease at a higher level (systolic blood pressure >160 mm Hg, cholesterol > 6.5 mmol, BMI kg/m$^2$ > 30) or two risk factors at a lower level (systolic blood pressure >140 mm Hg, diastolic blood pressure >90 mm Hg, BMI > 25 kg/m$^2$ cholesterol > 5.2 mmol, current smoking, family history of ischaemic heart disease).

2. **Population group (n=50)**
   Patients were recruited by random selection from age/sex stratified lists generated by the practice computer (with age strata of 18–33, 34–49, 50–64, and 65–80 years). The rationale for this population group was: (a) to provide the whole range of nutrients and foods in the population to more accurately assess the validation characteristics of the assessment tools—high risk patients alone would be likely to have a narrower, more extreme range; (b) potentially so that dietary assessment tools in general practice could be used to monitor the population; and (c) that population health promotion was included in the health promotion guidelines of the GP contract at the time the study was designed.

ASSessment INSTRUMENTS
The instruments were chosen after a review of all those available that had been designed to be used in the UK setting. A range of formats were selected to provide comparative validation, varying from quick assessment tools to more formal instruments, and corresponding to a range of uses in clinical care and research (see appendix 1).

DESCRIPTION OF ORIGIN, DEVELOPMENT, AND USE OF TOOLS

Brief assessment tools

**HEA1** (“What do you eat?”) This is a self completion (or nurses completion) food portions questionnaire that was developed by the Health Education Authority in Oxford (HEA), takes about 10 minutes to complete and uses portion sizes and food groups. It asks the subjects to estimate the number of portions eaten in a “normal” day, grouped according to foods types/groups—starchy foods (potatoes, pasta, rice bread, cereal, crackers) fruit and vegetables, meat or equivalent, dairy, fatty/snack foods. Portion size is not necessarily equivalent to normal serving size so subjects have to make this calculation for themselves. This tool was field tested but not previously validated.

**HEA2** During piloting, some patients reported difficulty saying what they ate on a “normal” day because days are not all the same, and thus we developed HEA2. HEA2 was identical in format to HEA1, but asked patients to assess intake for an average day or week. This tool allows foods not eaten every day to be put in a week column.

**HEA3** This tool was similar to HEA2, but separates portion sizes and food frequency. Some patients reported at the beginning of the main study, that they found it difficult to do the simultaneous portions/serving and food frequency calculation for HEA1/2. HEA3 was thus developed after the main study began, and separates estimation of food frequency and serving size. The original version of this tool omitted a dairy section; the agreement between HEA3 and HEA2 when the dairy section was omitted a dairy section; the agreement between HEA3 and HEA2 was very close (rank correlation 0.9): thus the tabulated results from HEA3 use the dairy results from HEA2 and are for 79 subjects.

**EPIC** This is a self completion food frequency questionnaire that was designed and validated for the research, specifically for the EPIC study to investigate diet/disease relations. Previous validation documented reasonable validation characteristics in a general population group.

**DINE** The Dietary Instrument for Nutritional Education (DINE) is a food frequency questionnaire designed for completion by a practice nurse, which takes about 20 minutes to complete, and from which a fat or fibre score can be generated, and an unsaturated fat score. Previous research in an occupational health
setting showed reasonable validation characteristics.

Nurse Questions We developed a structured nurse administered questionnaire based around type of foods normally eaten.

PGNC The Post Graduate Nutrition Centre (PGNC) at the Rowett Institute, Aberdeen, developed simple self completion scoring sheets for fat and fibre.

More interactive/intensive assessment tools

24 hour recall A standard proforma developed in conjunction with nutritionists and dietitians (BM, RT, SW, HW) was used to document recall of foods eaten in the previous 24 hours. This is a commonly used tool for research purposes, and takes 30–40 minutes to administer.

Checklist For seven days the subjects ticks for a page long checklist what foods they have eaten that day. This tools showed reasonable validation characteristics in a general population group.

Weighed record (the relative “standard”) A seven day weighed record was chosen as an accepted relative standard and to allow correct classification of patients into thirds of the distribution for major food components. Patients were given written instructions and demonstrations in how to use Soehnle 800100 digital scales and asked to weigh all food and drink consumed during seven days, and provide manufacturers data for foods where possible.

DESIGN CONSIDERATIONS: MINIMISING LEARNING ORDER, AND BALANCING SEASONAL DRIFT AND TEST-TEST INTERACTION EFFECTS

The flow diagram (fig 1) shows the design of the study, and when each tool was administered. To minimise the learning effect, assessments tools that required more interaction with the health professional (24 hour recall) or that involved the patient having to make a detailed inventory of their foods (the checklist)—likely to have the most learning effect—were retained as a later block. Within each block the order of administration of tools was randomised. To assess the possible learning effect that taking part in a study with multiple assessments might have on the performance of instruments, HEA2 was completed both at the beginning and end of the study. To minimise the effect of seasonal drift the whole study was completed within a three month period for each person, and to limit test-test interaction no tool was administered sooner than 10 days after the previous tool.

ASSESSING THE VALIDITY AND VARIABILITY OF THE STANDARD

Repeatability study Most validation studies do not assess the accuracy or variability of the standard. This is crucial in assessing validation data as agreement between a test tool and the relative standard can be no better than the agreement of the standard with itself. To assess the errors or variability in the standard, in a separate study sample 29 patients completed two weighed records six weeks apart (the average time between administering of a test tool and standard for the main study). Both DINE and HEA3 were also administered before completing each weighed record to assess the reliability of these simple tools. The data comparing DINE and HEA3 with a weighed record from the repeatability study were combined with the main study results for comparison of test with standard, to improve the precision of the estimates for these tools.

Total energy intake (TEI)/basal metabolic rate (BMR) ratio People taking part in dietary assessment studies under-report food intake. To document the extent of this under-reporting BMR was measured: patients with a TEI/BMR ratio of less than 1.2 were classified as under-reporters—that is people whose reported energy intake was incompatible with energy expenditure of free living subjects.

Para amino benzoic acid (PABA) urine collection There should be good agreement between documented protein or nitrogen intake from a valid standard and urinary nitrogen excretion measured from seven days of complete urine collections. A week of urine collections was not feasible—it was most difficult to get subjects to agree to compete urine collection. Subjects agreed to perform two 24 hour urine collections with estimation of PABA for completeness, thus allowing some estimation of the relative validity of the standard. Urine containing more than 85% of $3 \times 80$ mg PABA tablets was regarded as “complete”.

Coding data Each food item from the questionnaire or weighed record was coded (RT, DW) using the Royal Society of Chemistry data base, and converted to nutrient values. Although DINE was not originally designed to be coded in this way, given the widespread use of this instrument, it was also coded to facilitate comparison with other tools.

ANALYSIS

Data were entered and analysed using SPSS for Windows. As some of the variables were
skewed, non-parametric measures and statistics were used. Differences are presented as medians/interquartile range, and correlations as rank correlation (Spearman $r$). The novel presentation of results by important foods groups is justified as health care professionals are likely to assess and give counselling based on foods and food types rather than food components. To allow comparison of performance by each test tool across the range of both food groups and nutrients, percentage differences compared with the standard are reported: that is, $(\text{test value} - \text{standard value})/\text{standard value} \times 100$.

Thus, the main summary statistic for agreement between test and relative standard is the median percentage difference (and interquartile range for the difference). To assess whether the median difference/interquartile range is a fair summary of agreement across the range of values, correlations are quoted from a plot of the difference between test and standard and the mean of the two (Bland-Altman plot). Correlations are also quoted from a simple scatter plot to allow estimation of the linear association and the ability of the test tool to correctly rank subjects, and also to allow comparison with previous validation studies. Thus, a test tool performing well compared with the standard will have low median difference, narrow interquartile range of the difference, high rank correlation from a scatter plot, and low rank correlation from a Bland-Altman plot. As the agreement of the test tool with standard is never likely to be better than the agreement of the standard with itself, for reference, the comparable statistics from the standard-standard repeatability study are quoted in the same table as the test-standard results.

Predictors of under-reporting of energy intake were assessed using logistic regression. Terms were included in the model by forward selection if they were significant—and remained significant—at the 5% level (using the likelihood ratio $\chi^2$ test).

**Results**

**THE STUDY POPULATION**

There was no dietary information available for those not taking part: however, the socio-demographic characteristics of those agreeing to take part in the main study compared with those who did not agree are shown in table 1. Although those participating in the population group had fewer men, more social class I/II and fewer age less than 50 years compared with the high risk group, the participating group as a whole had very similar sociodemographic characteristics to those not participating and were not dissimilar from the national population. The nutritional characteristics (mean (standard deviation)) taken from the weighed record of the study population were as follows:

<table>
<thead>
<tr>
<th></th>
<th>Male sex</th>
<th>Social class I/II</th>
<th>Age &lt;50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population group participants (n=50)</td>
<td>19(38)</td>
<td>18(36)</td>
<td>28(56)</td>
</tr>
<tr>
<td>non-participants (n=196)</td>
<td>92(47)</td>
<td>12(24)</td>
<td>126(64)</td>
</tr>
<tr>
<td>High risk group participants</td>
<td>34(56)</td>
<td>13(21)</td>
<td>32(52)</td>
</tr>
<tr>
<td>non-participants</td>
<td>39(46)</td>
<td>8(25)</td>
<td>41(49)</td>
</tr>
<tr>
<td>Total participants (n=111)</td>
<td>53(48)</td>
<td>31(28)</td>
<td>60(54)</td>
</tr>
<tr>
<td>non-participants (n=280)</td>
<td>131(46)</td>
<td>20(81)</td>
<td>167(60)</td>
</tr>
</tbody>
</table>

Non-participants: patients who refused to take part (61%) or no response/not on phone/not suitable (39%).

**KEY POINTS**

- Most patients requiring dietary assessment and counselling are managed and assessed in general practice.
- The cost effectiveness of dietary counselling by nurses is limited by time spent on assessment.
- There has been no validation of appropriate instruments using health professionals in a general practice setting.
- Brief assessment instruments perform well enough for clinical work and research, and as well or better than more time consuming instruments.

**Table 1 Characteristics of participants compared with non-participants (number(%))**

<table>
<thead>
<tr>
<th>Group</th>
<th>Male sex</th>
<th>Social class I/II</th>
<th>Age &lt;50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population group participants (n=50)</td>
<td>19(38)</td>
<td>18(36)</td>
<td>28(56)</td>
</tr>
<tr>
<td>non-participants (n=196)</td>
<td>92(47)</td>
<td>12(24)</td>
<td>126(64)</td>
</tr>
<tr>
<td>High risk group participants</td>
<td>34(56)</td>
<td>13(21)</td>
<td>32(52)</td>
</tr>
<tr>
<td>non-participants</td>
<td>39(46)</td>
<td>8(25)</td>
<td>41(49)</td>
</tr>
<tr>
<td>Total participants (n=111)</td>
<td>53(48)</td>
<td>31(28)</td>
<td>60(54)</td>
</tr>
<tr>
<td>non-participants (n=280)</td>
<td>131(46)</td>
<td>20(81)</td>
<td>167(60)</td>
</tr>
</tbody>
</table>

Non-participants: patients who refused to take part (61%) or no response/not on phone/not suitable (39%).
Table 2 Percentage energy from fat and from saturated fat. Comparison of test tools with weighed record, and for comparison agreement of weighed record with itself (from reliability study). Percentage difference (median, interquartile range) and correlations (Spearman r from scatter plot and Bland-Altman plot quoted)

<table>
<thead>
<tr>
<th>Test Tool</th>
<th>% Difference (Median)</th>
<th>Bland-Altman r</th>
<th>Scatter plot r</th>
<th>% Difference (Median)</th>
<th>Bland-Altman r</th>
<th>Scatter plot r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighed record</td>
<td>2.1 (–0.1, 1.4)</td>
<td>–0.10</td>
<td>0.74</td>
<td>4.6 (–9.8, 30.1)</td>
<td>–0.02</td>
<td>0.58</td>
</tr>
<tr>
<td>Checklist</td>
<td>–6.4 (–15.4, 2.2)</td>
<td>–0.17</td>
<td>0.47</td>
<td>–6.7 (–18.8, 9.5)</td>
<td>–0.04</td>
<td>0.50</td>
</tr>
<tr>
<td>Epic</td>
<td>–1.9 (–14.1, 10.8)</td>
<td>0.14</td>
<td>0.50</td>
<td>–1.7 (–14.4, 12.2)</td>
<td>0.21</td>
<td>0.64</td>
</tr>
<tr>
<td>Structured Nurse Questions</td>
<td>5.0 (–5.5, 18.6)</td>
<td>0.14</td>
<td>0.46</td>
<td>16 (0.2, 34.1)</td>
<td>0.33</td>
<td>0.55</td>
</tr>
<tr>
<td>24 hour recall</td>
<td>3.6 (–12.2, 15.0)</td>
<td>0.38</td>
<td>0.45</td>
<td>1.0 (–19.1, 19.9)</td>
<td>0.51</td>
<td>0.45</td>
</tr>
<tr>
<td>HEA1 (food groups; portions per day)</td>
<td>4.7 (–11.4, 20.0)</td>
<td>0.39</td>
<td>0.35</td>
<td>10.5 (–13.5, 34.3)</td>
<td>0.53</td>
<td>0.37</td>
</tr>
<tr>
<td>HEA2 (food groups; portions per week)</td>
<td>–6 (–24.2, 2.8)</td>
<td>0.15</td>
<td>0.49</td>
<td>–5.0 (–15.8, 7.3)</td>
<td>0.1</td>
<td>0.60</td>
</tr>
<tr>
<td>HEA3 (frequency, portion size separately)</td>
<td>–7.8 (–17.8, 6.3)</td>
<td>0.22</td>
<td>0.51</td>
<td>16.9 (–10.5, 39.1)</td>
<td>0.46</td>
<td>0.59</td>
</tr>
<tr>
<td>DINE</td>
<td>–0.4 (–7.4, 15.3)</td>
<td>–0.01</td>
<td>0.38</td>
<td>4.6 (–6.5, 23.0)</td>
<td>0.52</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Table 3 Non-starch polysaccharide (NSP) and vitamin C. Comparison of test tools with weighed record, and for comparison agreement of weighed record with itself (from reliability study). Percentage difference (median, interquartile range) and correlations (Spearman r from scatter plot and Bland-Altman plot quoted)

<table>
<thead>
<tr>
<th>Test Tool</th>
<th>Median % Difference</th>
<th>Bland-Altman r</th>
<th>Scatter plot r</th>
<th>Median % Difference</th>
<th>Bland-Altman r</th>
<th>Scatter plot r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighed record</td>
<td>–9.6 (–24.0, 6.2)</td>
<td>0.10</td>
<td>0.78</td>
<td>8.5 (–20.5, 46.7)</td>
<td>–0.04</td>
<td>0.83</td>
</tr>
<tr>
<td>Checklist</td>
<td>11.8 (–8.9, 29.0)</td>
<td>0.18</td>
<td>0.65</td>
<td>34.8 (–6.4, 94.2)</td>
<td>0.31</td>
<td>0.56</td>
</tr>
<tr>
<td>Epic</td>
<td>26 (1.8, 57)</td>
<td>0.17</td>
<td>0.38</td>
<td>54.6 (13.2, 137.8)</td>
<td>0.44</td>
<td>0.56</td>
</tr>
<tr>
<td>Structured Nurse Questions</td>
<td>3.9 (–18.8, 26.5)</td>
<td>0.27</td>
<td>0.57</td>
<td>–4.4 (–24.0, 47.2)</td>
<td>0.11</td>
<td>0.52</td>
</tr>
<tr>
<td>24 hour recall</td>
<td>–9.4 (–31.7, 23.7)</td>
<td>0.24</td>
<td>0.41</td>
<td>–8.6 (–49.1, 52.0)</td>
<td>0.32</td>
<td>0.58</td>
</tr>
<tr>
<td>HEA1 (food groups; portions per day)</td>
<td>0.3 (–24.3, 29.3)</td>
<td>0.19</td>
<td>0.29</td>
<td>46.1 (7.9, 114.6)</td>
<td>0.28</td>
<td>0.63</td>
</tr>
<tr>
<td>HEA2 (food groups; portions per week)</td>
<td>9.1 (–12.5, 40.4)</td>
<td>0.25</td>
<td>0.53</td>
<td>34.6 (–6.9, 101.9)</td>
<td>0.13</td>
<td>0.52</td>
</tr>
<tr>
<td>HEA3 (food groups number of times per week; portion size separately)</td>
<td>–20.9 (–43.9, 0.6)</td>
<td>0.03</td>
<td>0.42</td>
<td>32.1 (–17.4, 80.3)</td>
<td>0.07</td>
<td>0.46</td>
</tr>
<tr>
<td>DINE</td>
<td>–17.9 (–34.8, 14.6)</td>
<td>–0.23</td>
<td>0.47</td>
<td>–40.4 (–56.1, 14.4)</td>
<td>–0.84</td>
<td>0.45</td>
</tr>
</tbody>
</table>
and the simple instruments devised by the PGNC “score” fat and fibre, and these scores agree well with the weighed record: the Spearman rank correlation with the weighed record for fat, saturated fat, and NSP are 0.53, 0.49, 0.56 respectively for the PGNC tools, and 0.51, 0.47, and 0.46 respectively for DINE.

CHECKS ON THE STANDARD

**TEI/BMR ratio**

Forty per cent of subjects had a TEI/BMR ratio of less than 1.2—that is, under-reporting total energy intake—a similar to other dietary studies. If under-reporting is a limiting factor in the agreement between test and standard, then excluding these subjects should improve the agreement between test and standard. Excluding under-reporters did not change levels of agreement; thus for HEA2 the % energy from fat rank correlation was 0.50, % difference median = 6.3 (interquartile range −20.8 to −0.53); for NSP rank correlation 0.44, % difference 9.9 (interquartile range −16.2 to 38.4). This suggests that, in primary care settings, subjects who under-report for the standard probably under-report for other assessment instruments, and that under-reporting is not confined just to the lower end of the range of intake. The obese (BMI ≥30) were much more likely to be “under-reporters” compared with the non-obese (60% versus 30% respectively, \( \chi^2 \) 10.9, p<0.001); no other factors (age, sex, class, study group) predicted under-reporting once obesity was controlled for.

**TWENTY FOUR HOUR NITROGEN EXCRETION, PROTEIN INTAKE, AND PABA**

There was a good correlation between the average nitrogen excretion for the mean of two 24 hour urine collections and protein intake from the weighed record (Spearman rank correlation 0.57), which provides evidence supporting the validity of the weighed record. However, 50% of subjects had incomplete urine collection as judged by the PABA estimation, which probably reflects the inconvenience and difficulty of collecting 24 hour urine specimens in free living subjects. Excluding those subjects where PABA collection was incomplete did not improve agreement between urinary nitrogen and protein intake (rank correlation 0.47).

**LEARNING EFFECT**

If a significant learning effect was present then the agreement between HEA2 and the standard would be improved when repeated at the end of the study—with a narrower interquartile range for the difference and higher rank correlation for the scatter plot. There was no evidence of any better agreement comparing HEA2 on the second occasion with the standard for % energy from fat (median % difference −7.5 (IQR −19.2 to 7.7) scatter plot rank correlation 0.46) or for NSP (median % difference −2.9 (−19.9 to 29.7) scatter plot rank correlation 0.42).

**Discussion**

This study provides evidence that the relative validity of brief dietary assessment tools used in general practice is as good as—or better than—more complicated instruments validated in other settings. This is the first comparative dietary validation study performed by a practice nurse among subjects likely to be assessed in general practice, and thus is likely to provide the most generalisable results for primary care of any validation study to date.

**LIMITATIONS OF THE STUDY**

Before discussing the main results, the general and specific limitations of validation studies must be considered.

**Reporting bias, the use of biomarkers, and precision of the “standard”**

Validation studies have many difficulties inherent in trying to assess diet, including recall bias and the change in behaviour that occurs when diet is measured, and thus absolute or “gold” standards of food intake are unrealistic. Evidence that subjects do change “reported” behaviour—in particular that they are likely to under-report energy intake—is confirmed by the low TEI/BMR ratio for many subjects in this, and in most other validation studies. Thus, “absolute” levels of intake are likely to be systematically under-reported at the group level, and this study suggests that this is much more likely if subjects are obese. Expense and logistic difficulties prevented using doubly labelled water as an energy biomarker, although the 24 hour urinary nitrogen collection did provide evidence for the relative validity of the standard, and quoting percentage energy rather than absolute intake probably limits the
bias caused by under-reporting.\textsuperscript{14} Furthermore, many of the committed participants prepared to accept the conditions of this study still did not submit complete urine collections. Better assessment of “under-reporting”—including the assessment of differential bias—will require more freely available, less inconvenient biomarkers that do not require multiple urine collections.\textsuperscript{13, 14} Most validation studies report modest agreement with the validation standard, but interpreting this is difficult as the errors/variance in the standard are not reported. We report the errors, or variation, in the standard to facilitate interpretation of the results.

Generaliseability
Potential threats to generaliseability are that “patients” who agree to the conditions of validation studies may be unusual, and in this study were drawn from a single practice base. However, we have shown that those participating do not differ significantly in socio-demographic characteristics from those not participating, and are not dissimilar from the national population. Furthermore the generaliseability of the dietary assessment was maximised by using written proformas throughout for each dietary assessment tool.

Utility of results for clinicians and researchers
The generalisability of dietary assessment to the clinical setting is limited in most studies as most studies do not report results for foods, which are likely to be the basis for counselling in health care settings. We report the agreement by food groups. Thus, with some reservations, this study is likely to give a reasonable estimate of the relative validity of available tools against a “relative” standard, which should generalise to both research and clinical work in primary care.

Validation characteristics of the test instruments
The agreement between test and standard for % energy from fat, % energy from saturated fat, NSP, was reasonable for both simple and more complex instruments, particularly when compared with the estimated agreement of the weighed record with itself. As primary health care professionals are likely to counsel using food groups and types of food, it is important that test instruments provide validation data for foods: as with food constituents there was moderate agreement between test instruments and the standard for both fruit and vegetables and for starchy foods.

Time and complexity of assessment
An important finding is that simple self-completed tools, which require little health professional time, do as well or better than simple nurse administered tools, and often as well or better than much more complicated instruments designed for greater accuracy in research. Whether simple tools would do as well for the range of micronutrients likely to be important for health is unclear; however, close agreement for micronutrients would be unlikely particularly for important foods rarely eaten. Thus, at least for major food groups or food components, there is little advantage, either for research or clinical purposes, in performing complicated time consuming assessment. Given that nurse time is the main determinant of cost effectiveness of intervention,\textsuperscript{6} reducing nurses time for dietary assessment should make dietary intervention—that is, including both assessment and counselling—much more cost effective.

Which simple instruments perform best?
Simple assessment instruments based on number of portions and food groups—“what you eat in the normal week”—perform well across food constituents and for foods (HEA2 or HEA3). The performance for most nutrients and also for food groups, comparable to the performance of other common clinical measurement,\textsuperscript{18, 19} suggests that using such tools will make more efficient use of nurses time. Some patients found calculating the number of weekly portions difficult (HEA2), and thus for some patients separating the elements of this calculation may be preferable (HEA3). The advantage of these tools is that is easy to provide simple advice about the balance of food types compared with recommended guidelines; the disadvantage is that in the current format there is no ready feedback to patients about the foods that contribute the most to fat/saturated fat (a computerised version of the tools linking foods to food constituents has not yet been developed). Approaches based on the “normal day” (HEA1, 24 hour recall) are not likely to provide accurate classification of the person within the distribution, but provide reasonable estimate at the group level for most components and foods. This suggests a less useful role in individual counselling but a useful role in population monitoring or in calibration studies. The extensively used DINE tool performed moderately well as in the previous validation study although less well than the food group tools, and has the advantage that scores can be easily calculated as the basis of counselling advice. In its current form the major disadvantage of DINE is the probable underestimation of fruit and vegetable consumption. The other disadvantage is that is intended for the nurses to complete. Given that it is not more accurate than the other simple tools it is very difficult to justify nurses using their time to complete it, particularly as this extra time may not improve results\textsuperscript{10}; however, we have shown from the repeatability study that the DINE tool can be self-administered with no great loss in accuracy.

Conclusion
Simple tools designed for dietary assessment by practice nurses perform as well as much more complicated assessment instruments and are comparable to the performance of the standard itself. The variability of standards should be assessed in validation studies by test-retest methods to allow interpretation of test-standard agreement. Under-reporting is
Appendix 1

Instruments used

<table>
<thead>
<tr>
<th>Time period (days)</th>
<th>Format</th>
<th>Time (min)</th>
<th>Data entry time (computer in/output)</th>
<th>Counselling guidance in manual format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Checklist</td>
<td>self completion; tick foods eaten from daily list</td>
<td>30–60</td>
<td>10–20</td>
<td>no</td>
</tr>
<tr>
<td>EPIC</td>
<td>self completion; food frequency</td>
<td>30–40</td>
<td>20</td>
<td>no</td>
</tr>
<tr>
<td>Nurse Questions</td>
<td>interview</td>
<td>20</td>
<td>5–10</td>
<td>no</td>
</tr>
<tr>
<td>24 hour recall</td>
<td>interview</td>
<td>30</td>
<td>10</td>
<td>no</td>
</tr>
<tr>
<td>HEA1</td>
<td>self completion; calculate number of portions by food groups</td>
<td>10–15</td>
<td>5–10</td>
<td>yes — based on type of food and food groups</td>
</tr>
<tr>
<td>HEA2</td>
<td>self completion; calculate number of portions by food groups</td>
<td>10–15</td>
<td>5–10</td>
<td>yes — based on type of food and food groups</td>
</tr>
<tr>
<td>HEA3</td>
<td>self completion; number of portions by food groups but no portion calculation needed</td>
<td>10–15</td>
<td>5–10</td>
<td>yes — based on type of food and food groups</td>
</tr>
<tr>
<td>PGNCE</td>
<td>self completion</td>
<td>5–10</td>
<td>5–10</td>
<td>yes — based on fat/fibre/P:S scores</td>
</tr>
<tr>
<td>DINE</td>
<td>nurse completion (self possible)</td>
<td>10–20</td>
<td>5–10</td>
<td>yes — based on fat/fibre/P:S scores</td>
</tr>
</tbody>
</table>

common and more likely in obese subjects. There is little justification for most clinical or research purposes, where major food components or food groups are being assessed, in using time consuming dietary assessment methods, as simple self completion methods perform equally well, and well enough to provide useful discrimination for both clinical work and research.

Funding: Dr Little and Sister Barnett are supported by the Wellcome trust. This work was supported by the Wellcome trust, the South and West Regional Health Authority NHS research and development directorate, the Health Education Authority, and the Rank Prize Fund.

We are grateful for the help of Drs Nigel Dickson, Ian Williamson, Stephen Morgan, Simon Griffin, Marion Dyer, Patrick Terry, and Fiona Bradley for their help recruiting patients; to Dr David Rowe, Jo Kelly and Angela Rounsford for biochemical estimations and assessment of biomarkers.