Validity of Dietary Assessment in Athletes: A Systematic Review

Louise Capling1,6* (acap7726@uni.sydney.edu.au), Kathryn L. Beck2 (k.l.beck@massey.ac.nz), Janelle A. Gifford1 (janelle.gifford@sydney.edu.au), Gary Slater3 (gslater@usc.edu.au), Victoria M. Flood1,4,5 (vicki.flood@sydney.edu.au), Helen O'Connor1,4 (helen.oconnor@sydney.edu.au)

1. The University of Sydney, Discipline of Exercise and Sport Science, Faculty of Health Sciences, Lidcombe, NSW, Australia
2. School of Sport Exercise and Nutrition, College of Health, Massey University, Auckland, New Zealand
3. School of Health and Sport Sciences, University of the Sunshine Coast, Maroochydore, Qld, Australia
4. The University of Sydney, Charles Perkins Centre, Camperdown, NSW, Australia
5. Western Sydney Local Health District, Westmead, NSW, Australia
6. Queensland Academy of Sport, Sport Performance Innovation and Knowledge Excellence, Brisbane, Qld, Australia

* Correspondence: acap7726@uni.sydney.edu.au; Tel.: +61 427 179 080

Received: date; Accepted: date; Published: date

Abstract: Dietary assessment methods recognized as appropriate for the general population are usually applied in a similar manner to athletes, despite knowledge that sport-specific factors can complicate assessment and impact accuracy in unique ways. As dietary assessment methods are used extensively within the field of sports nutrition, there is concern the validity of methodologies have not undergone more rigorous evaluation in this unique population sub-group. The purpose was to
systematically review studies comparing two or more methods of dietary assessment, including dietary intake measured against biomarkers or reference measures of energy expenditure, in athletes.

Six electronic databases were searched for English-language, full-text articles published from January 1980 until June 2016. The search strategy combined the following keywords: diet, nutrition assessment, athlete and validity; where the following outcomes are reported but not limited to: energy intake, macro and/or micronutrient intake, food intake, nutritional adequacy, diet quality, or nutritional status. Meta-analysis was performed on studies with sufficient methodological similarity, with between-group standardized mean differences (or effect size) and 95% confidence intervals (CI) calculated. Of the 1624 studies identified, 18 were eligible for inclusion. Studies comparing self-reported energy intake (EI) to energy expenditure assessed via doubly labelled water were grouped for comparison (n=11) and demonstrated mean EI was under-estimated by 19% (-2793 ± 1134 kJ/d).

Meta-analysis revealed a large pooled effect size of -1.006 (95% CI: -1.3 to -0.7; p<0.001). The remaining studies (n=7) compared a new dietary tool or instrument to a reference method(s) (e.g. food record, 24-h dietary recall, biomarker) as part of a validation study. This systematic review revealed there are limited robust studies evaluating dietary assessment methods in athletes. Existing literature demonstrates substantial variability between methods, with under and misreporting of intake frequently observed. There is a clear need for careful validation of dietary assessment methods, including emerging technical innovations, among athlete populations.

Keywords: dietary assessment; food record; FFQ; biomarker; doubly labeled water; energy intake; validity; athletes; sports nutrition.

1. Introduction

Adequate dietary intake is important for normal growth and development, maintaining health and well-being, reducing the risk of illness and injury, and optimizing sports performance [1]. Individual
dietary requirements are influenced by a range of factors such as age, gender, body mass, stature, and growth and development needs (for child or adolescent athletes) [1,2]. In addition, a range of sport-specific factors such as type of sport, training volume and intensity also influence dietary requirements which are not static due to periodization of training load across days, weeks or months of a competitive season [1,2]. Athletes have special nutrition needs which usually encompass a higher energy requirement to account for greater energy expenditure, increased protein and carbohydrate requirements to support lean mass accrual and/or maintenance, and glycogen stores, respectively as well as an increased requirement for certain micronutrients (e.g. iron, calcium, sodium) [1,2,3].

Dietary assessment is routinely undertaken by nutrition professionals to evaluate whether an individual is achieving specific health and/or sports nutrition targets [4]. However, accurate dietary assessment of athletes is complex due to the influence of sport-specific factors as well as the widespread use of rapidly evolving sports foods and supplements. Dietary assessment can be expensive (especially for large numbers of athletes or teams), time consuming and unfortunately due to these constraints, undertaken less frequently. As a result, athletes with inadequate or poor dietary intake may not be readily identified.

Numerous studies have reported on dietary intake of athletes from a wide range of sports [5-9]; with many comparing self-reported intake to calculated energy requirements based on general recommendations and/or sports nutrition guidelines [10-14]. However, the discrepancy between self-reported intake and estimated energy expenditure has been reported in a number of studies involving male and female athletes, across a range of age-groups and a variety of sports [15-17]; with substantial differences (11–44 %) observed [18,19].
Dietary assessment methodology recognized as suitable for non-athletes are usually applied in a similar manner to athletes, although specific factors such as training status, competition level and nutrition-associated beliefs and dietary practices could influence aspects of the dietary assessment process [19]. Many retrospective methods that rely on self-reporting of intake (e.g. food record, 24-h dietary recall, diet history) are susceptible to measurement error including conscious or sub-conscious exclusion of foods consumed [4,19-21]; and possible change in usual food intake or dietary patterns due to the nature of the dietary intervention itself [22].

By far, the most common dietary assessment method applied in sports nutrition research and practice is the food record (FR); where all food and drinks consumed are recorded by the participant for a specified number of days (i.e. 3-7 days) [4]. However, it has been widely reported that self-reported FR under-estimates intake as noted in a number of population groups including adolescents and children [23], obese individuals [24], and athletes [18,25,26]. Despite the FR being regarded a ‘gold standard’ in dietary assessment, it places substantial burden on participants to document information truthfully and accurately, as well as reliance on investigators to code the data correctly using appropriate databases [19,27-29]. Additional challenges facing an athlete cohort include increased burden from recording large food intakes, frequent eating occasions, potentially irregular meal patterns, difficulty estimating large portion sizes, and contribution of sports foods and supplements [19].

Recent technological developments involving the use of food photography or electronic images, wearable cameras, and various on-line tools and applications have been shown to improve participant compliance by reducing the burden of recording and enhance the accuracy of data recorded [30-33]. There is a need for high quality research to assist in identifying dietary
methodologies, including emerging technologies, which are valid, as well as feasible, for use in this
unique population sub-group.

Despite a clear need for careful validation of dietary assessment methods in athletes, there has
been relatively little attention directed to the applicability and validity of dietary assessment methods
in this population sub-group. The purpose of this systematic review was to evaluate studies
comparing nutrition intake from two or more dietary assessment methods, including intake
measured against biomarkers or reference measures of energy expenditure, in athletes.

2. Methods

2.1 Search Strategy

A systematic literature search was conducted of articles published between January 1980 to June
2016 to identify studies comparing the nutrition intake assessed by two or more methods of dietary
assessment, including a measure of intake against dietary biomarkers or reference measures of
energy expenditure, in an athlete cohort. Databases searched were AUSPORT MED (via Informit
Online), CINAHL (via EBSCO), EMBASE (via EBSCO), MEDLINE (via OvidSP), SPORTDiscus (via
EBSCO) and the Web of Science. Ongoing electronic monitoring was established to assess the
inclusion of eligible recent publications.

The search strategy combined the following keywords: diet (e.g. energy intake, nutritional
status, nutritional requirement, food group); nutrition assessment (e.g. food frequency questionnaire,
food record, diet score, diet survey, energy expenditure, biomarkers); athlete (e.g. elite athlete, team
sport, collegiate, Olympic athlete); and validity (e.g. under-reporting, measurement error). In
addition, the reference lists of all retrieved papers and relevant reviews were manually searched for
eligible papers. Following the search, we completed a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [34] informed systematic review process.

### 2.2 Selection of Studies

Study participants had to be ‘athletes’ (i.e. amateur, collegiate, tertiary, professional, national or international calibre) as defined by the study authors. Included studies needed to validate at least one dietary assessment method (e.g. food frequency questionnaire, 24-h dietary recall, food record) against at least one other dietary assessment method or a biological (blood or urine) measure for nutrient status (e.g. 24-hr nitrogen excretion, antioxidant concentrations). In addition, studies which compared reported energy intake to measured energy expenditure (e.g. doubly labelled water technique, activity records) were also included. Descriptive, cross-sectional or longitudinal study designs were eligible for inclusion. Studies including participants with a mean age lower than 16 years of age, papers published in a language other than English, case studies, reviews, abstracts and theses were excluded (Supplementary Material 1). After eliminating duplicates, one reviewer (LC) screened the search results against the eligibility criteria. References that could not be eliminated by title and abstract were retrieved in full and subsequently assessed against the eligibility criteria in duplicate (KB, GS) prior to inclusion.

### 2.3 Data Extraction and Conversions

Data relating to the manuscript, namely author(s), date of publication and country where the study was conducted were extracted. The institution country of the first author was used if the study country was not described in the paper. Data extracted from each publication included: participant characteristics (i.e. age, sex, sport and athlete calibre), participant anthropometric (i.e. stature, body mass) or body composition parameters (e.g. percent fat), and dietary assessment methodology (e.g.
food record, food frequency questionnaire, 24-h dietary recall). In addition, information on reported
energy and macronutrient intake, resting energy expenditure, total energy expenditure, and key
statistical comparisons between dietary assessment methods and correlation coefficients, were
extracted in conjunction with main study findings. Where a publication contained insufficient
information, attempts were made to contact the study author(s) to obtain missing details. To aid
comparison between the included papers, anthropometric parameters reported in imperial units (e.g.
feet, inches, pounds) were converted to metric units (e.g. kg, cm) (1 kg = 2.2 pounds; 1 cm = 0.3937
inches). Data reported in calories (e.g. energy intake, energy expenditure) were converted to
kilojoules (1 Cal = 4.184 kJ). Where reported, macronutrient values were converted from grams or
grams per kilogram body weight to per cent of mean energy intake using Atwater factors (i.e. protein
17 kJ/g or 4 Cal/g; fat 37 kJ/g or 9 Cal/g; carbohydrate 16 kJ/g or 4 Cal/g) [35]. Extracted data were
presented as mean and standard deviation (SD) where reported. Weighted means were calculated
for age, anthropometric variables and the differences between reported energy intake and measured
energy expenditure.

2.4 Assessment of Methodological Quality

The methodological quality of all included studies were independently assessed by three
researchers (LC reviewed all papers; KB and JG shared the evaluation of the other studies) using a
modified assessment scale created by Downs and Black [36]. Using the scale, 18 of the 27 criteria that
logically applied to the study designs were retained, while items 4, 8, 9, 13, 14, 17, 21, 23, 24 were
deemed not relevant to the included studies and were excluded. Three additional items from a
nutrition-specific quality criteria checklist were incorporated [37]; specifically item 8.6 Was clinical
significance as well as statistical significance reported?; item 9 Are conclusions supported by results with
biases and limitations taken into consideration?; and item 10 Is bias due to study’s funding or sponsorship unlikely? Where relevant, other items from the Academy of Nutrition and Dietetics quality criteria checklist [37] (i.e. items 2.1, 2.3, 3.4, 5.2, 7.2, 7.3, 7.4, 7.5 and 7.6) were considered concurrently to the Downs and Black [36] criteria to provide reviewer clarification when assessing quality ratings. Two variables from a third checklist described by Serra-Majem and colleagues [38] were applied specifically to validation studies to evaluate the validity of the dietary methodology used and suitability of the method for an athlete cohort. The specific variables applied to assess the quality of the validation study were statistics to assess validity (i.e. to support Downs and Black [36] item 18: Were the statistical tests used to assess the main outcomes appropriate?); and to assess whether the sample size was adequate to assess validity (i.e. to support Downs and Black [36] item 27: Did the study have sufficient power to detect a clinically significant effect where the probability value for a difference being due to chance is less than 5%? (Supplementary material 2). Each reviewer checked for internal (intra-rater) validity across items for each included publication. Disagreements in ratings were resolved by discussion and through adjudication with two separate researchers (HO, VF). No further sub-analysis was undertaken on the basis of methodological quality and none of the studies were eliminated.

2.5 Meta-Analysis

Studies which involved the comparison of a dietary measure (i.e. food record) to energy expenditure as measured by DLW were sufficiently similar methodologically to enable meta-analysis to be performed. To quantitatively compare differences between reported and measured energy intake in the papers using DLW, the between-group standardised mean difference, or effect size (ES) and 95 % confidence interval (CI) was calculated. The extracted data (i.e. mean energy intake, mean energy expenditure, SD and sample size) was transferred into Comprehensive Meta-Analysis (CMA)
version 2 software (Biostat, 2005, Englewood, IL, USA) in order to calculate the ES, standard error, variance and 95% CIs. A forest plot was generated to display the ES and 95% CIs from each study and the pooled estimate determined whereby a small ES was > 0.2, a medium ES > 0.5 and a large ES > 0.8 [39]. A positive ES indicated an effect favouring energy measured by DLW while a negative ES indicated an effect favouring reported EI.

3. Results

3.1 Identification and Selection of Studies

The initial literature search identified 1624 potentially relevant papers. After removal of duplicates and elimination of papers based on title and abstract, 42 full text publications were reviewed in duplicate against the inclusion criteria. A manual search of reference lists identified an additional nine papers. Thirty-three papers were excluded because they did not meet the inclusion criteria, leaving 18 full-text articles for assessment (Figure 1).

Figure 1 PRISMA flow diagram of identification, screening and selection process for included articles

3.2 Demographic and Anthropometric Characteristics

Of the 18 studies included, 11 compared reported energy intake (EI) against energy expenditure (TEE) as measured by DLW (TEE\textsubscript{DLW}) [21,22,40-48]. The remaining seven papers [29,49-54] compared
a new dietary tool \((n=3)\) or FFQ or modified FFQ \((n=4)\) to a reference method(s) (i.e. food record, 24-h dietary recall, direct observation, biomarker) as part of a validation study. Of the 18 included studies, three compared one or more dietary assessment method(s) to a biomarker (i.e. urine \(n=2\), blood \(n=1\)) \([45,50,54]\).

The total number of study participants was 683 (range 7-156 participants); with smaller sample sizes noted in papers comparing energy intake to DLW (range 7-19 participants). A range of sports were represented most of which were ‘mixed sports’ (6 studies; \(n=461\) participants) with the remainder reporting on athletes from ‘aerobic’ or ‘endurance sports’ (e.g. long distance running, swimming, rowing, triathlon, cross country skiing) (9 studies; \(n=184\)), ‘team sports’ (e.g. soccer, basketball) (2 studies; \(n=26\)), and ‘skill’ or ‘aesthetic sports’ (1 paper; \(n=12\) ballet dancers). The mean age was 21.8 ± 2.6 years (range 16-30.4 years). Two studies did not report participant age however it was assumed the participants were older than 16 years of age due to recruitment from tertiary institutions \([41,53]\). More than half of the studies were in male athletes (55 % participants). Athletic caliber included competitive tertiary, national, international and professional level athletes. The studies were conducted in Europe \((n=6)\), North America \((n=6)\), and across the Asia Pacific and African regions \((n=6)\) (Table 1).

Physique characteristics of the athletes are summarized in Table 1. The weighted mean body mass and stature were 65.3 ± 7.2 kg and 173.3 ± 6.1 cm, respectively. One paper did not report mean stature of participants \([50]\). Seven studies reported percentage body fat with mean levels ranging from 7.1 ± 2.5 to 22.8 ± 5.1 % (Table 1).
Table 1 Participant characteristics of athletes from included studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Age (mean), years</th>
<th>Body Mass Index (mean)</th>
<th>Height (mean), cm</th>
<th>Weight (mean), kg</th>
<th>Sport</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1</td>
<td>22</td>
<td>25</td>
<td>180</td>
<td>75</td>
<td>Football</td>
<td></td>
</tr>
<tr>
<td>Study 2</td>
<td>24</td>
<td>26</td>
<td>182</td>
<td>80</td>
<td>Basketball</td>
<td></td>
</tr>
<tr>
<td>Study 3</td>
<td>20</td>
<td>24</td>
<td>178</td>
<td>70</td>
<td>Soccer</td>
<td></td>
</tr>
</tbody>
</table>

**3.3 Studies Comparing Reported Energy Intake to Energy Expenditure as Measured by DLW**

Eleven studies assessed the difference between reported mean EI and mean TEE as measured by DLW (Figure 2). The most common method for measuring EI was by weighed or estimated food records (FR) over 4-10 days (mean 6.5 days) duration (Table 2). Where reported (n=5), the mean macronutrient intake (as percent of energy intake) was 56.5, 14.9 and 26.8% for carbohydrate, protein and fat, respectively. One paper reported mean protein intake only (12.6% of energy intake) [45].
Energy expenditure was measured for a mean of 8.4 days (range 6-14 days) by the DLW method (Table 2). Details on the training phase or competitive season of the participants were reported by six studies [21,22,40,42,43,47]. Six studies also recorded physical activity levels during the assessment period either by self-reporting the number of training hours via an activity record (n=5) [22,41,45,46,48], or as measured by ActiGraph™ activity monitor (n=1) [42].

Eight of the eleven studies measured dietary intake for the entire DLW measurement period [21,22,40-45,47]; while three studies measured intake for only 4-5 days [43,44,48]. Most studies (n=7) reported a stable body mass during the testing period, however four studies adjusted energy for changes noted in body mass during the testing period [43-45], or for poor dietary records as a result of participant non-compliance [22] (Table 2).

Resting metabolic rate (RMR) was measured via indirect calorimetry in four studies [40,43,46,47]; four reported RMR based on prediction calculations (i.e. Schofield, Cunningham or FAO/WHO/UNU equations) [21,42,44,48]; while three studies did not report RMR [22,41,45] (Table 2).
3.3.1 Mean difference between EI and TEE: reporting bias

Compared to DLW, all studies reported a lower mean EI of 19% (range 0.4 - 36%) with a weighted mean difference of $-2793 \pm 1134$ kJ/d between measures (Table 3). The difference was significant in seven studies [21,40-42,44,45,47] while smaller, non-significant differences were reported in the remaining studies [22,43,46,48]. For example, Hill and Davies [44] found a very strong...
inverse correlation of difference between TEE and self-reported EI, and the means of the two
measurements (r=-0.93, p<0.01) in elite female lightweight rowers. When mean EI was adjusted for
change in body mass (1.2 ± 1.2 kg) over the measurement period, the correlation of difference
between the mean adjusted TEE and self-reported EI was also significant (r=-0.93, p<0.01). The limits
of agreement (LOA) (95% CI ± 2 SD) for the mean adjusted TEE and self-reported EI were – 17 619
to 8134 kJ. The authors also noted that reporting bias increased in those with higher TEE (i.e. TEE >
3000 kcal/d; or 12 552 kJ/d). In contrast, Schulz and colleagues [46] found no relationship between
self-reported EI and TEE in female distance runners and reported a mean difference of - 925 ± 2301
kJ/d. Similarly, Sjodin and colleagues [48] noted a close match between reported EI and TEE (i.e. mean
difference of - 100 ± 1900 kJ/d) in eight cross-country skiers (Table 2).

Koehler and colleagues [45] aimed to assess the validity of a food and activity record compared
to reference methods such as DLW, 24-h urea nitrogen excretion and indirect calorimetry (i.e. to
determine energy expenditure during an incremental running or cycling test). They found EI and
TEE as measured by DLW were only weakly correlated, however following the exclusion of two
implausible results (where EI < 1.39 x REE) [55] a significant positive correlation was observed (r=0.69,
p<0.05). A proportional bias towards over-estimating low and under-estimating high energy intakes
(p<0.01) was also noted with very wide LOA (95% ± 2 SD) between EI and TEE of - 5736 and 4912
kJ/d (Table 2).
3.4 Meta-Analysis

Studies which involved the comparison of a dietary measure (i.e. self-reported or weighted FR) to TEE assessed by the DLW technique were homogenous enough to conduct a meta-analysis (n=11). Relevant data (i.e. mean EI, SD and sample size) were used to calculate the between-group standardised mean difference, or effect size (ES) and 95% confidence intervals (CIs). The meta-analysis revealed a large pooled ES of -1.006 (95% CI -1.3 to -0.7; p=0.00) where reported EI was significantly below TEE measured by DLW (mean -2477 kJ/d) (Figure 3).

Figure 3 Meta-analysis of the pooled effect of self-reported EI compared to TEE measured by DLW

3.5 Studies Comparing Reported Dietary Intake by Two or More Methods of Dietary Assessment

Of the seven studies comparing reported intake by two or more methods of dietary assessment, four compared energy and/or macronutrient intake [49,51-53], while the remaining three studies compared the intake of a specific nutrient (i.e. antioxidants, calcium, or protein) [29,50,54]. Only two studies validated the test method (i.e. FFQ, modified FFQ) for micronutrients and food groups [51,53] (Table 4).
Four studies investigated the validity of a new self-administered, semi-quantitative or quantitative food questionnaire (i.e. FFQ or modified FFQ) in an athlete cohort. The dietary reference methods used to evaluate these new instruments included a FR (n=3) [29,50,51] or multiple pass 24-h DR (n=1) [53]. Braakhuis and colleagues [50] compared a blood biomarker of antioxidant capacity (i.e. ferric-reducing ability of plasma [FRAP]) with antioxidant intakes recorded from a 7-d weighed FR and FFQ. The remaining three studies aimed to validate a new dietary assessment instrument (e.g. web-based 24-h DR, virtual interface tool). The dietary reference methods used to evaluate these new instruments included a FR (n=1) [52]; multiple pass 24-h DR (n=2) [49,54]; and/or direct observation (n=1) [49]. Wardenaar and colleagues [54] also compared self-reported protein intake as measured by 24-h DR to urinary nitrogen excretion (Table 4).
3.5.1 Reported mean energy intake

Six of the seven studies reported mean EI and/or compared intake between dietary assessment methods (Table 4). Sunami and colleagues [53] compared mean EI of college athletes from a variety of sports from a FFQ (138 items) with 3-d non-consecutive 24-h DR and found the FFQ underestimated EI by 9% (males) and 10% (females). In comparison, Fogelholm and Lahti-Koski [51] found close agreement on a group level of mean EI of mixed sport athletes from a food use questionnaire (FUQ) compared to a 7-d weighed FR (95% CI -1.7, 0.1 FUQ1, 95% CI -0.1, 1.7 FUQ2); however, individual agreement was weak. Scoffier and colleagues [52] found no difference between EI measured by a virtual interface dietary assessment tool compared to a 1-d FR in two groups (i.e. weight sports, and other mixed sports) of adolescent athletes. Baker and colleagues [49] compared reported energy and macronutrient intakes measured by a web-based, 24-h DR tool (DATA) to a 24-h DR (interview) and observation (by registered dietitian), and found no significant difference between the measures. They found good relative validity for group level comparisons but large variations of individual dietary intake estimates, especially in athletes with higher energy and macronutrient intakes (Table 4).

3.5.2 Reported mean macronutrient intake
Where reported (4 of 7 studies), the mean macronutrient intake (as percent of EI) was 53.1, 15.3 and 28.5% for carbohydrate, protein and fat, respectively (Table 4). An additional study reported mean protein intake only (11.9% of EI) [54]. Wardenaar and colleagues [54] reported a mean difference between 24-h DR and 24-h nitrogen excretion (25.5 ± 21.3% difference, or 31.7 ± 30 g/d; p<0.001). Under-reporting of protein was related to the amount of protein intake (r=−0.20; 95% CI -0.46, 0.09) such that under-reporting was greater at a higher protein compared to a lower protein intake. In contrast, Koehler and colleagues [45] found protein intake in triathletes was strongly correlated to 24-h urinary nitrogen excretion (r=0.81; SEE=0.34 g/kg/d), however considerable individual variation was observed between the two methods with very wide limits of agreement (-0.65 to 0.67 g/kg/d).

3.5.3 Other nutrients, food groups and dietary patterns reported

Braakhuis and colleagues [50] found a modest correlation between energy-adjusted estimates of total antioxidant intakes in competitive rowers as measured by a quantitative FFQ (70 items) compared to 7-d weighed FR (r=0.38, 90% CI ± 0.14), and only a small correlation between the FFQ and plasma biomarker (r=0.28, 90% CI ± 0.15). However, the authors noted a trend to over or under-estimate antioxidant intakes at low and high intake levels (42% over and 73% under-estimation, respectively). Ward and colleagues [29] compared mean calcium intake as measured by a self-administered calcium checklist to a 6-d FR (822 ± 331 mg/d and 823 ± 387 mg/d, respectively) and observed no significant difference between the two methods (Table 4).

In addition to comparing energy and macronutrient intake between the test method (i.e. FUQ1 and FUQ2) and a 7-d weighed FR, Folgelholm and Lahti-Koski [51] reported on selected micronutrients (i.e. thiamin, vitamin C, calcium, magnesium, iron and zinc). They found the FUQ1
over-estimated reported intake of vitamin C, calcium, magnesium, iron and zinc; while the FUQ2 did not differ in mean values for the selected micronutrients compared to a 7-d FR. Most of the correlations by food group were reported above $r = 0.24$, with the exception of ‘vegetable oils’, ‘cream’, ‘milk’, ‘pork’, ‘beef’, and ‘poultry’ [51].

One study validated the test dietary assessment method (i.e. FFQ) compared to a 3-d non-consecutive 24-h DR for 35 nutrients and 19 food groups [53]. While the difference in measures for the majority of nutrients was within ± 20 %, the largest difference was observed for retinol (i.e. 77 % for males and 32% for females). The authors found lower differences between measures for ‘cereals’ (1 % in males, 10 % in females), ‘vegetables’ (9 % in males, 4 % in females) and ‘fungi’ (5 % in males, 9 % in females) food items; while greater differences were noted for ‘sugar’ (- 94 % in males, - 96 % in females), ‘beverages’ (-45 % males, - 54 % females), and ‘seasonings and spices’ (- 65 % males, - 72 % females) food items. Overall, under and over-estimation of food group intakes was more prevalent than that of nutrient intakes [53] (Table 4).

### 3.6 Evaluation of Methodological Quality

The methodological quality of all included studies ($n=18$) were evaluated with quality ratings determined as fair to moderate for the majority of studies (Table 5). All papers stated the aim or hypothesis, described the main outcomes to be measured, described the main findings, reported clinical and statistical significance, and accounted for participant drop outs. All but two studies provided estimates of the random variability [43,46]; reported on data dredging [44,48]; and all but two reported sources of funding [44,45]. However, the poorest ratings were for the following items: “Were actual probabilities reported?” (Item 7); “Were the subjects asked to participate representative of the entire population?” (Item 10); “Were the participating subjects representative of the population they were
recruited from?” (Item 11); and “Were the subjects recruited over the same time period?” (Item 18) Only two studies [49,54] reported whether an attempt was made to blind those measuring the main outcomes of the intervention (i.e. Item 12) (Table 5).

Table 5 Evaluation of methodological quality of included studies

4. Discussion

This is the first systematic review to have evaluated studies comparing two or more dietary assessment methods, including measuring intake against biomarkers or reference measures of energy expenditure, in athletes. Only 18 papers were published over the past 37 years, confirming the limited literature in this area. Most papers (n=11) focused on self-reported EI compared to TEE as measured by the DLW technique [21,22,40-48]. The remaining studies (n=7) compared a dietary assessment method to a reference method(s) (e.g. food record, 24-h DR, direct observation, biomarker) as part of a validation study [29,49-54]. Dietary assessment methods recognized as appropriate for the general population are usually applied in a similar manner to athlete groups, despite the knowledge that athlete-specific factors can complicate the assessment and impact accuracy in unique ways. As dietary assessment is used extensively in both sports nutrition clinical and research settings, it is a concern the validity of the methodologies used have not undergone more rigorous evaluation. There
is a clear need for high quality research to be undertaken to identify dietary assessment methodologies which are valid, as well as feasible for use, in an athlete population.

4.1 Studies Comparing EI to TEE as Measured by DLW

Overall, the studies (n=11) comparing self-reported EI to TEE DLW found mean EI was underestimated by 19% (range 0.4-36%) which is comparable to differences observed in studies in other populations [18-20,25,26,56,57]. Three DLW studies [22,45,48] reported close agreement (<5% difference between measures); which can be partially explained as participants were provided with a test diet for the duration of the DLW collection and a participant was excluded due to an inexplicable marked difference in agreement between measures [22]. Koehler and colleagues [45] adjusted mean EI following removal of participants (n=2) for loss of body mass (>3%) or due to implausible food records as determined by a cut-off value (i.e. EI<1.39*RMR) [55]. While Sjodin and colleagues [48] noted a close agreement between mean EI and TEE DLW over 7 days (r=0.96; p=0.0001); no relationship was observed when authors compared EI from separate 24-h periods, indicating athletes were not in energy balance during shorter periods of time.

The remainder of the DLW studies (n=8) reported a greater mismatch between self-reported EI and TEE DLW (12-34% difference) attributed primarily to misreporting [21,40-44,46,47]. A variety of explanations have been proposed for lower reported EI when compared to TEE DLW [21,25,40,46]. Hill and Davies [18] suggest factors such as body size, perception of body image, restrained eating, gender, socioeconomic status, motivation, social expectations and the nature of the testing environment itself, play a role in misreporting. Nutrition-related beliefs and dietary practices of athletes present additional challenges with reporting intake accurately in this population sub-group. In addition, factors such as high energy requirements, frequency of snacking, eating away from the
home environment, applicability of ‘standard’ portion sizes, and the wide use of commercial sports foods, drinks and supplements can make it more difficult to quantify food intake [8,19].

4.1.1 Methodological Issues

The DLW technique has long been considered a ‘gold standard’ for measuring TEE and to validate self-reported EI [20]. While the DLW technique is considered accurate to 1%, with a coefficient of variation of 2-12%, depending on the loading dose and length of sampling period [20,58]; potential sources of error should be considered when measuring in free-living conditions including the influence of day-to-day variation of EI and TEE [58]. The optimal duration for measuring DLW is between 2-3 biological half-lives of the isotopes [59]; although a shorter period (i.e. 8-12 days) has been suggested for athlete cohorts due to higher rates of water turnover due to regular physical activity [59]. A potential limitation is the DLW measurement period of most of the included studies (\(n=8\)) may have been too short (i.e. 6-7 days duration) [21,40-42,45,46-48].

4.1.2 Assessment of Dietary Intake

Typically dietary intake is measured in DLW studies to assist in the calculation of the respiratory quotient (RQ); a required variable in calorimetry equations [60]. This calculation was reported in the methodology of only four of the included studies [22,40,46,47]. In non-athletes, a 3–4d FR is considered valid for assessment of RQ and EI for groups [61]; however dietary intake of athletes may be more variable due to the day-to-day variation in the energy cost of training. It has been suggested that recording intake for 3-7 days is a reasonable compromise between scientific rigor and practicality when estimating dietary intake of athletes to capture habitual intake and high variability in day to day energy expenditure [19,28,62]. Hence, for athletes, ideally intake would be recorded over the
entire DLW measurement period to capture this variation. As such, a further limitation is that three
studies did not record intake for the same duration as the DLW measurement [43,44,48].

In all of the included DLW studies, dietary intake was determined by weighed or estimated FR
(recorded for 4-10 days). The documentation of intake via FR is considered practical and is widely
used in both clinical and research settings, however the accuracy of data is reliant on subjective
recording of intake, level of participant motivation, and possible fatigue from recording for longer
durations [18,19,25]. Recording intake itself can modify usual eating behavior [46,63]. Investigators
are responsible for improving compliance by providing clear, specific instructions to ensure records
provide sufficient detail. Nine studies provided detail about instructions provided to participants
and review process of dietary records upon collection [21,40-44,46-48]. Only one study supervised
the weighed FR over a limited observation period in order to improve recording accuracy [48]; while
another provided a test diet to all participants [22].

Given the potential for misreporting during the recording period, the plausibility of self-
reported EI can be compared to a pre-determined EI:BMR value [55]. Two DLW studies applied a
cut-off value to exclude implausible dietary records [44,45], and one of the validation studies [50].
However, when applying a cut-off value to assess self-reported EI of athletes, it is important to have
a valid assessment of the energy demands of the athletes being assessed, as the range in TEE can be
wide and is reliant on the type of sport, training phase and intensity [19,56,64].

4.1.3 Variability of Intake and Expenditure in Athletes

Although the DLW studies investigated energy balance or misreporting of intake, this was not
the primary aim of all included papers. A number of the studies (n=4) aimed to determine TEE of a
group of athletes (i.e. swimmers, soccer players, light weight rowers, distance runners) and compare
to recommendations [21,40,44,46]; with the secondary outcome to evaluate the agreement between reported EI and TEE_{DLW}. The primary aim of the remaining papers (n=7) was to compare the difference between EI and TEE_{DLW} [22,41,45,47]; assess energy balance in a group of athletes (i.e. endurance runners, cross-country skiers) [42,48]; and assess the validity of a 4d FR in measuring EI of female classical ballet dancers [43].

While DLW is highly regarded as an accurate method for the validation of reported EI [64], a major limitation is that the technique is unable to calculate energy turnover on a daily basis. In athletes, TEE (and EI) can fluctuate substantially from day to day [65,66]. For example, Bradley and colleagues [65] found EI of rugby union players lowest when TEE was highest early in the training week, with EI increasing in preparation for competition on the weekend. Similarly, Brown and colleagues [66] noted differences in EI, macronutrient intake and energy balance of dancers between weekdays and weekend days. These recent publications highlight the day-to-day variability in energy demands of athletes over a relatively short period of time which is unable to be assessed by DLW.

4.1.4 Influence of Body Mass, Body Image and Energy Demands

In most studies there was no significant change in body mass during the assessment period [21,22,40,41,45-48] indicating the difference between measures was likely due to misreporting.

Conversely, Fudge and colleagues [42] attributed the difference noted between EI and TEE_{DLW} (-13%) in elite Kenyan runners to under-eating resulting in a negative energy balance during a period of intense training. Four studies adjusted mean TEE_{DLW} based on changes in body mass during the measurement period [22,43,44,46]. Previous research has indicated the degree of misreporting increases with increasing body mass, specifically adiposity [67,68] and BMI (body mass index) [69].

Others suggest that misreporting is independent of adiposity but linked to restrained eating, and
body image [18,70-72]. In the current review, Edwards and colleagues [41] observed a relationship between EI and body mass (r=-0.74) (i.e. heavier women reported a lower relative intake). However, Hill and Davies [43] found no relationship between the extent of misreporting and percent body fat of female ballet dancers (r=0.11). Similarly, Silva and colleagues [47] found no relationship between EI and body mass or composition in male and female basketball players.

It appears the trend for misreporting increases with increasing TEE, particularly for individuals with high energy needs [19,57]. Possible reasons include increased burden from reporting large volumes of food and frequent eating occasions resulting in food or drink items being omitted consciously or unconsciously. In the current review, a negative correlation was observed between self-reported EI and TEE\textsubscript{DLW} in some studies (r=-0.854, p<0.01) (i.e. the higher the expenditure, the lower the reported intake) [40,41]. While there was almost no difference between self-reported EI and TEE\textsubscript{DLW} in a group of triathletes, the data indicated a strong proportional bias towards under-estimating high energy intakes (p<0.01) [45]. Similarly, Hill and Davies [44] reported a bias to misreporting with increasing EE (i.e. TEE > 3000 kcal/d; or 12 552 kJ/d) in female lightweight rowers. This trend is supported by findings in non-athletes where the magnitude of misreporting intake increases with increasing TEE [57,73,74].

4.2 Studies Comparing Dietary Intake by Two or More Methods of Dietary Assessment

The current review identified a limited number of studies (n=7) which evaluated the validity of a novel dietary assessment tool or instrument compared to one or more reference method(s) [29,49-53] and/or a biomarker in an athlete population [50,54]. However, due to limited consistency in the pairing of study methods, it was difficult to make firm conclusions on the relative strength of the different methodologies used to evaluate dietary intake in athletes. Despite most studies reporting
acceptable validity for group level comparisons [49,51,54], individual agreement was not as strong, especially for athletes with high energy intakes [49,50,54]. A variety of statistical tests were used to assess study validity (e.g. paired t-tests, correlation coefficients, Bland-Altman plots), however most studies used correlation coefficients [50,53] or intra-class correlation (ICC) to assess reproducibility [29,49]. Correlation coefficients can be misleading, as they measure the relationship between methods, rather than the agreement between them [75]. Only one study validated the test method for food groups [53], while the other studies evaluated energy and/or nutrient intakes [29,49-52,54]

4.2.1 Dietary Reference Methods

Three of the included studies determined dietary intake by (weighed or estimated) FR (6-7 days) [29,50,51]; while another recorded intake for a single day only [52]. Research suggests that a longer recording duration is necessary to assess habitual intake and account for day-to-day variation of intake and TEE in athletes, with a 7-d FR two to three times less variable than a 1-d FR [28]. Other reference methods which rely on self-reporting (e.g. 24-h DR, FFQ) are also prone to measurement error associated with recall bias and awareness of portion size [19,53,76]. Three studies used the multiple pass 24-h DR [49,53,54] and/or direct observation [49] to evaluate a new tool or instrument. While 24-h DR has the advantage of low participant burden, data may not be representative of the usual diet unless the recalls are repeated a number of times [4,19]. In addition, Wardenaar and colleagues [54] suggest dietary data should be collected within a four week period to provide the best insight into the accuracy of the multiple 24-h DR in athletes, due to the micro-cycle and periodization of training (i.e. training programs with a variable workload and volume throughout the year). It has been previously suggested that the 24-h DR is prone to underestimate EI, and that caution should be taken assessing data involving high intakes [4,19,77]. In general, all dietary assessment methods are
influenced by errors of precision and validity [4,67]. As such, it is possible the validation studies included in the current review failed to detect true reporting bias, if both the new and established dietary assessment instruments have correlated error [20], particularly when applied to an athlete population. Researchers are encouraged to consider the relative strengths and weaknesses of the dietary reference method(s) selected, in addition to consider participant burden, cost, and validity in the study population [53].

4.2.2 Evaluation Using Biomarkers

Two of the included validation studies used a biomarker to assess the accuracy of reported intake [50,54]. Wardenaar and colleagues [54] found a multiple 24-h DR acceptable for ranking protein intake compared to the reference urinary nitrogen biomarker (r=0.65). However, they noted the ‘standard’ portion sizes used in the 24-h DR may not be representative for athletes with high intakes of protein and therefore a potential source of error [54]. Braakhuis and colleagues [50] found only a weak association between the plasma antioxidant biomarker (FRAP) and FFQ (r=0.28). However relying on blood biomarkers to validate a nutrition questionnaire can be problematic as there is no single marker for antioxidant intake [50]. Unfortunately, the use of independent biomarkers to assess the accuracy of dietary intake is limited to energy intake or a specific nutrient only (i.e. sodium, nitrogen). Koehler and colleagues [45] indicate that their research first to validate dietary protein intake against 24-h urinary nitrogen excretion in an athlete population. They found good agreement between 24-h nitrogen excretion and dietary protein intake (r=0.81) which is consistent with results documented in non-athlete populations [56,64,76]. More recently, energy availability (EA) (i.e. EA=EI-ExEE) has been linked to biochemical indices such as cortisol, insulin,
growth hormone, IGF-1, leptin and thyroid hormones [45,78]. These present potential use as biomarkers in future dietary assessment studies.

4.2.3 Nutrients, Food Groups and Dietary Patterns

Micronutrient intake and food group intake is less frequently reported in dietary validation studies than energy and macronutrient intakes [38]. One paper evaluated calcium intake in female college athletes from a range of sports (e.g. basketball, cross-country) using a 6-d weighed FR to validate a self-administered calcium checklist [29]. Mean calcium estimates did not differ between measures, however longer periods of recording may be required to estimate intake of key micronutrients such as iron and calcium (i.e. up to 11 recording days) [19,61,62,77]. In another study, Braakhuis and colleagues [50] found an FFQ was valid for estimating antioxidant intake in elite rowers.

Two studies reported on a range of selected micronutrients [51,53] as part of validating test methods (i.e. FUQ, FFQ, respectively). These same studies are the only studies in athletes to validate dietary intake at the food group level. Sunami and colleagues [53] found a FFQ was useful for assessing habitual dietary intake of college athletes for vitamin C, calcium, vegetables, fruits, and milk and dairy products. While Fogelholm and Lahti-Kosti [51] found close agreement of EI between the FR and FUQs, with most food group correlations above r=0.24, except for ‘vegetable oils’, ‘other fats’, ‘standard milk’, ‘pork’, ‘beef’ and ‘poultry’.

More recently, nutrition epidemiology has progressed from examining nutrients per se to exploring the relationship of dietary patterns or diet quality on health outcomes [79,80]. However to date, there has been limited evaluation of diet quality or dietary patterns in athletes [8,81-83]. Dietary patterns, including the specific timing of intake over the day, have been shown to enhance health,
training and performance outcomes [1,83] and should be addressed in future studies. Acknowledging
there are challenges with assessing intake accurately, it may prove beneficial in understanding
dietary patterns or diet quality of athletes to help identify athletes who may benefit from dietary
input, and provide a platform to educate individuals about dietary choices for optimal health and
sports performance.

4.3 Qualitative Assessment of Methodological Quality

Overall the studies included in this review (n=18) were of fair to moderate quality. Poor ratings
were generally noted for items which evaluated external validity and internal validity. For example,
selection criteria (i.e. inclusion and exclusion criteria) and representativeness of the participants was
usually not adequately described. Ten studies included an adequate description of participant
characteristics (i.e. item 3) [21,22,40-42,46,47,49,50,54]; while only six studies [21,29,42,50,51,54]
clearly identified the source and how participants were recruited (i.e. item 10) and/or stated the
proportion who were asked agreed to participate in the study (i.e. item 11). Potential confounding
factors such as supplement use, maintenance of body mass, or maintenance of usual physical activity
levels were often not discussed. The description of and adjustment for confounding factors (i.e. items
4 and 19, respectively) were either unable to be determined or deemed not relevant for most of the
DLW papers [21,22,40,41,44-48] due to the nature of methodology involving DLW studies. Finally,
attempts made to blind investigators to the main outcomes of the intervention were poorly rated for
all but two studies [49,54]. Despite known challenges with reporting dietary intake accurately using
FR; monitoring of and compliance with self-reported intake data was not always considered in the
included studies. Twelve studies indicated that the expertise of qualified dietitians were involved in
the collection and/or analysis of dietary data [22,41-44,46-51,54].
5. Limitations, Strengths and Future Directions

One of the key strengths of the current review includes the extensive, systematic search and evaluation of the literature which has compared two or more dietary assessment methods, including measuring intake against dietary biomarkers or reference measures of energy expenditure in athletes. The results indicate there are limited robust studies evaluating the validity of dietary assessment methods in athletes. The caliber of participants could have influenced study findings and relevance for an elite athlete population. For example, a number of studies involved participants from university sporting clubs or teams [22,29,41,43,49,53], while two studies did not provide sufficient detail about the recruitment source or participant athletic caliber [45,52]. The relatively small sample size in all but two of the eighteen papers could influence the power required to detect meaningful differences. For validation studies, Serra-Majem and colleagues [38] suggest including a sample size greater than 100 participants, or more than 50 participants when using a biomarker. However, it is not uncommon for sports science research to include a relatively small population sample which could make it difficult to detect significant change [84]. There may also be challenges in recruiting a sample that is powerful enough and representative of high level athletes due to reticence of athletes to commit to research studies perceived as time consuming and detracting from training or competition commitments. A further consideration is that the isotope deuterium used in DLW studies is expensive which could influence sample size due to cost and research feasibility.

It has been suggested that combining dietary assessment methods may enhance the accuracy of assessing dietary intake [85]. For example, Rumbold and colleagues [86] found a combination of FR and 24-h DRs were effective when quantifying EI in adolescent netball players (i.e. 4.2% difference between measures indicating a slight bias towards over-reporting). In a similar study protocol, Briggs...
and colleagues [87] compared the accuracy of a combined dietary data collection method (i.e. weighed FR and 24-h DR) to observed EI of adolescent male soccer players. The results showed systematic under-reporting of intake compared to observed intake (p<0.01), but the bias was small (i.e. degree of random error between dietary methods 3.1%). These publications suggest a combined dietary data collection method may provide an effective technique when quantifying energy intake in athletes.

Finally, emerging image-assisted technical innovations such as wearable cameras, handheld devices and mobile telephone technology have been shown to improve participant compliance by reducing the burden of recording and enhance the accuracy of data recorded [30-33]. Results from a systematic review by Gemming and colleagues [32] indicate that images enhance self-report by revealing unreported foods and identify misreporting errors not captured by traditional assessment methods alone. The use of an image-based FR deployed via a mobile application has recently been shown to be a positive tool for dietary monitoring and potential influence on dietary habits and behaviors [88].

6. Conclusions

Adequate dietary intake is crucial for the maintenance of health and optimizing performance outcomes of athletes. However, there are unique challenges in assessing intake, including the day-to-day variation of expenditure and wide use of commercial sports foods and supplements, which can challenge the accuracy of dietary assessment methods used in athletes. Results from this review suggest self-reported food records may not be a suitable assessment method for quantifying energy expenditure, particularly for weight conscious athletes or athletes with high energy demands. Existing literature demonstrates substantial variability between dietary assessment methods, with
under and misreporting of intake frequently observed. As such, there is a clear need for careful validation of dietary assessment methods with emerging technical innovations likely to show promise as they may assist with portion quantification, reduce the burden of collection and problem with missing foods, among athlete populations.

Supplementary Materials: The following are available online at www.mdpi.com/link, Supplementary material 1: Search strategy and inclusion criteria; Supplementary material 2: Quality ratings guide.

Acknowledgments: The authors would like to acknowledge funding support by the Queensland Academy of Sport’s Sport Performance Innovation and Knowledge Excellence unit, Brisbane, Australia, which assisted in the preparation and submission of this manuscript.

The authors would also like to acknowledge the technical expertise of Ms. Elaine Tam who assisted with the search of relevant databases and retrieval of articles from The University of Sydney Health Sciences library.

Author Contributions: The primary author (LC) and co-authors (HO, VF, JG, GS, KB) were responsible for the conception, design and planning of the manuscript; analysis and interpretation of the data; and drafting and subsequent revisions of the manuscript. Contributing authors were also involved in decisions involving final article inclusion (KB, GS, LC, HO), extraction of data (KB, LC), quality ratings (JG, KB, LC, HO, VF), and revision of the manuscript and final approval for publication (KB, JG, GS, VF, HO, LC).

Conflicts of Interest: The authors declare no conflict of interest.
References


9. Wardenaar, F.; Brinkmans, N.; Ceelen, I.; Van Rooji, B.; Mensink, M.; Witkamp, R. and De Vries, J. Macronutrient intakes in 553 Dutch elite and sub-elite endurance, team, and strength


food-frequency questionnaire to assess short-term antioxidant intake in athletes. *Int. J. Sport

51. Fogelholm, M. and Lahti-Koski, M. The validity of a food use questionnaire in assessing the

52. Scoffier, S.; Gernigon, C.; Billi, E. and d’Arripe-Longueville, F. Development and preliminary
validation of a new instrument to assess eating behaviors: The virtual self-service restaurant


Validation of a web-based, multiple 24-h recalls combined with nutritional supplement
intake questionnaires against nitrogen excretions to determine protein intake in Dutch elite

55. Goldberg, G.R.; Black, A.E.; Jebb, S.A.; Cole, T.J.; Murgatroyd, P.R.; Coward, W.A. and
Prentice, A.M. Critical evaluation of energy intake data using fundamental principles of


61. Basiotis, P.P.; Welsh S.O.; Cronin F.J.; Kelsay J.L. and Mertz W. Number of days of food intake records required to estimate individual and group nutrient intakes with defined confidence. J. Nutr. 1987, 117(9), 1638-41.


70. Heitmann, B.L. The influence of fatness, weight change, slimming history and other lifestyle variables on diet reporting in Danish men and women aged 35-65 years. *Int. J. Obes.* 1993, 17, 329-36.


