What are the maximum protein requirements of strength athletes?

A Systematic Review

"Dissertation submitted in accordance with requirements of University of Chester for the degree of Master of Science"

30th September 2008

Acknowledgements

I would like to acknowledge and thank Dr Stephen Fallows for his guidance and support when producing this review.

Abstract

BACKGROUND: Protein intakes above the recommended dietary allowance (RDA) for adult males have been suggested to be essential in accreting lean body mass, particularly in active individuals. Though, what is the maximum protein requirement of strength athletes in order to maximise their performance.

OBJECTIVE: A systematic review was conducted on all primary literature to establish the maximum protein requirements of strength athletes.

METHODS: A comprehensive search strategy involving searches of six electronic databases and 'grey' literature were conducted. The search was restricted to studies published after 1986 to the present day. All primary research literature that presented the effect of total dietary protein intake on lean body mass was included. Studies that met the inclusion criteria were assessed for methodological quality using the Downs and Black checklist.

RESULTS: 4 studies were identified that met the inclusion criteria, although only 3 studies met the quality assessment criteria; two randomised trials and one non-randomised trial. Statistically non-significant trends ($p>0.05$) deriving from muscle mass measurements, determined that the maximum protein requirement for strength athletes to be a moderate quantity of 1.4g/kg Bw/day. Similar results were shown in all three studies.

CONCLUSION: There is a sparsity of evidence and an inconsistency in the methodological designs between trials, regarding what the maximum protein requirement of strength athletes to be. Yet, it is likely to be a moderate protein intake, rather than a high protein intake.

Declaration

"This work is original and has not been submitted previously in support of a degree qualification or other course"

30/09/08

Contents

List of figures

List of tables

What are the maximum protein requirements of strength athletes? A Systematic Review

Background

Introduction

The protein requirements of strength athletes, individuals that engage in a regular resistance training programme to induce muscle mass gains, have been a topic of debate for many years, (Rasch & Pierson, 1962; Lemon, 1991). These athletes have long believed that their protein intakes must be much greater than for those who are sedentary or inactive, and indeed evidence supports this claim, (Lemon, Tarnopolsky, MacDougall & Atkinson, 1992; Tarnopolsky *et al*., 1992; American Dietetic Association, 2000). However, there is no additional allowance of protein intake for those involved in regular physical activity, above the current recommendations for men over 19 years of age. These recommendations all centre on 0.8 grams per kilogram of Bodyweight per day, (g/kg Bw/day), (Food & Nutrition Board, 1989; National Health & Welfare, 1990; Department of Health, 1991). But what quantity of protein can be recommended to strength athletes when considering their physical demands?

This systematic review reports the results of primary research evidence investigating the highest protein requirement of strength athletes, in that additional protein intake above this level will yield no further lean body mass gains.

Protein requirements:

Protein is an important part of our daily diet, required by the body for growth, maintenance and repair of cells, (Wilmore & Costill, 2004). It is recommended that 10-15% of total daily calorie intake should be made up of protein in a balanced diet, (Department of Health, 1991) or 0.8g/kg Bw/day, based on average body mass and allowing for an adequate margin of safety. This value was agreed following a consultation with Food and Agricultural Organisation (FAO) and World Health Organisation (WHO) representatives. However, this quantity has not been revised since 1989, (Food and Nutrition Board, 1989). Ultimately, this value derives from nitrogen balance studies which estimate the protein required for nitrogen equilibrium or for the optimal maintenance of the body. Although, this value presumes that the dietary protein is coming from a mixed diet containing a reasonable amount of good quality proteins, (Rodriguez & Garlick, 2008). So, this value may not be adequate enough for individuals consuming mixtures of poor quality proteins.

Excessive dietary protein is catabolised directly for energy or stored as fat, but excessive amounts can also be harmful to an individual's health, by possibly contributing to the demineralisation of bone and increasing urinary calcium loss, leading to bone loss, (Allen, Oddoye & Margen, 1979; Garn & Kangas, 1981) although this may not be the case, as increased dietary protein intakes usually accompany increases in phosphorus, which may minimise any such effect, (Hegsted, 1976; Flynn, 1985). Excessive dietary protein intakes also increase glomerular filtration rates, which may relate to an age-related decline in renal function, (Brenner, Meyer & Hostetter, 1982) although, this effect has shown to be greater with animal proteins than with vegetable proteins, (Wiseman *et al*., 1987).

Overall, there is no indication of any adverse health effects with excessive protein intakes, unless there are some pre-existing complicating factors such as abnormal kidney function, (Lemon, 1998).

Application to strength athletes

Protein has long been considered a key nutritional component for success in the sporting world from coaches of Olympians in ancient Greece to today's elite athletes, (Tipton & Wolfe, 2004). It is of particular importance to athletes participating in strength-based sports, (Lemon, 1998; Lemon, 2000; Tipton & Wolfe). Because protein is an essential element of skeletal muscle making up approximately 20% of its protein mass, (Wilmore & Costill, 2004) strength athletes consume a high amount of protein in dietary and supplement form in order to maximise their potential muscle mass gains. Though, these positive gains will only occur so long as there is a sufficient diet and training stimulus present, (Lemon et al., 1992; Tipton & Wolfe, 2001).

Protein requirements:

The current recommended dietary allowance (RDA) for protein does not recognise any increased protein need for physically active individuals, such as strength athletes, (Food Nutrition Board, 1989; Department of Health, 1991). However, this recommendation is likely to be incorrect, as it is based on data collected from physically inactive and minimally active individuals, (Lemon, 2000) so may not affect the protein requirements of strength athletes. A prominent theory supports the increased protein requirements for these athletes, in that there appears to be a direct relationship between protein requirements and the intensity and volume of training, (Laritcheva, Yalovaya, Shubin & Smirnov, 1978; Burke *et al*, 2001; Hoffman, Ratamess, Kang, Falvo & Faigenbaum, 2006) on the basis that an increase in intensity and volume of training may require an increased protein intake to

support muscle growth. As the intensity and volume of training are high in a strength athletes' training regime to support the highest potential muscle growth, protein intake above the RDA may be essential.

Nitrogen balance:

The underlying mechanism to the protein requirements of an individual depends on the metabolic demands of the body, in particular, the requirements of skeletal muscle mass per se. The concept of the 'nitrogen balance' explains these requirements with nitrogen being found in the amino acids in protein. The protein requirements of an individual are the amounts that balance all nitrogen losses (urine, faeces and sweat) to maintain nitrogen equilibrium and thus, to maintain skeletal muscle mass, (Millward, 2001). As previously stated, the RDA of protein to maintain a nitrogen balance status in sedentary individuals is 0.8g/kg Bw/day, (Department of Health, 1991) however, a strength athlete would logically need to be in a positive nitrogen status for accretion of lean body mass to occur. Therefore, it is safe to presume the highest protein requirement of a strength athlete will be above 0.8g/kg Bw/day, but to what extent?

Why maximise protein intake?

For a strength athlete, maintenance of skeletal muscle mass would not be considered a desirable objective. Instead, the protein requirements of these individuals would be an amount necessary for muscle growth to occur, possibly above 0.8g/kg Bw/day. Yet, strength athletes still consume diets very high in protein as high as 2-4g/kg Bw/day, (Tarnopolsky,

MacDougall & Atkinson, 1988; Steen, 1991) contrary to the scientific evidence suggesting that these diets are not necessary. This is largely due to recent evidence involving endurance exercise and not strength exercise, so the scientific data may be considered irrelevant, and that an athletes' own experimentation may have convinced them that high protein diets are advantageous. This underlying rationale cannot be dismissed since those who developed the RDA and subsequent revisions, did not address the dietary protein needs of athletes engaged in regular, rigorous training because the recommendation was designed for the general population, (Lemon, 1998).

Limitations associated with protein intake:

There is a protein intake above which no further muscle growth occurs, as additional protein ingested above this value would not contribute to muscle protein mass. This is due to the limitations of protein metabolism, in that the protein synthetic machinery of the body cannot process free amino acids within body tissue and the blood at a sufficient rate to synthesis tissue protein (lean body mass) up to a point, so the inevitable fate of the protein is that it is oxidised. This synthetic rate plateaus in sedentary individuals at around the RDA (0.8g/kg Bw/day), (Lemon, 1998) so logically strength athletes' synthetic rate must plateau at higher protein intakes.

Problems associated with identifying maximum protein requirements:

There is considerable controversy as to the protein requirements of these athletes deriving from nitrogen balance studies, due to indifferent findings in these research studies. Numerous studies found that no increases in protein intake are necessary, due to an increased efficiency in protein utilisation, (Butterfield & Calloway, 1984; Todd *et al*., 1984). This is achieved through the accommodation of muscle protein metabolism, as training ameliorates the response of the muscle to a bout of resistance training, (Phillips, Tipton, Ferrando & Wolfe, 1999; Phillips *et al*., 2002). Alternatively, Tome and Bos, (2000) stipulate that high protein intakes can result in continuous positive nitrogen balance from 1 to 3g/kg Bw/day, and Oddoye and Margen, (1979) found that a positive nitrogen balance was maintained for up to 50 days on a very high protein diet, up to three times the RDA with no adaptation evident, thus, muscle growth was still occurring. However, these indifferent findings are likely to derive from technical difficulties, such as unclear participant selection criteria, indifferent adaptation to protein intake, and an individual's current training stimulus. These factors cannot be controlled for entirely, however, this review will aim to maximise the considerations of these issues.

Rationale

This systematic review aims to identify the maximum protein requirements of strength athletes, but the reasons for its importance are stated below.

Rationale 1

It is deduced that higher protein requirements would be necessary for strength athletes to resistance train at higher intensities, (Laritcheva *et al*., 1978; Burke *et al*., 2001; Hoffman *et al*., 2006) which are conducive to stimulating growth hormone release and thus, muscle growth, (Felsing, Brasel & Cooper, 1992).

Relevance

Therefore, knowing the maximum protein requirements of strength athletes will optimise these athletes' physical performances'.

Rationale 2

Despite the scientific evidence indicating that carbohydrates and fats are the major exercise fuels, (Astrand & Rodahl, 1977) the majority of strength athletes have continued to consume high protein diets to excessive levels, at a point at which will cause protein oxidation and no further contribution to muscle mass accretion). Yet, these athletes are at

risk of developing health problems associated with these diets, (Allen, Oddoye & Margen, 1979; Garn & Kangas, 1981; Brenner, Meyer & Hostetter, 1982).

Relevance

Therefore, it would be beneficial for strength athletes to know their maximum protein requirements, to prevent the needless risk of health problems from occurring.

Rationale 3

As lifestyles have become more hectic in terms of longer working hours in recent years, (Boheim & Taylor, 2004) and perceived protein requirements are high, (Lemon, 1998) protein supplements in drink and snack forms are used to complement the habitual protein diet to meet these requirements. In effect, supplements act as meal replacements and are a great convenience.

Relevance

By identifying the maximum protein requirements of strength athletes, it would determine whether protein supplements are required on top of the athletes' habitual protein intake. For instance, if the protein requirements of these athletes were deemed low or nearer the RDA of 0.8g/kg Bw/day, then there would be little support for the use of supplements.

Rationale 4

Acquiring foods and supplements containing protein can carry a financial burden, especially with high protein diets. This becomes even more of an issue when considering that these strength athletes may be of an amateur level, and so may not have the disposable income necessary to purchase them.

Relevance

Identifying the maximum protein requirements of strength athletes, would determine whether high amounts of protein containing foods and supplements are required, for instance, if the maximum protein requirements were low, there would likely to be no associated financial burden on the athlete.

Aims/Objectives

The rationale forms the reasoning behind conducting the systematic review, though the primary objective of this review was to identify the maximum protein requirements of strength athletes, above which no further contribution to lean body mass per se would be evident.

The Review

The methods required to produce the review were developed in accordance with established standards for systematic literature reviews. The primary reference was Cochrane Collaboration, (Higgins & Green, 2008) an online library providing a reliable source of evidence through systematic reviews. Although, the evidence within the Cochrane library was not appropriate for this review as only primary research was included for data analysis, and that Cochrane reviews are mostly based on clinical trials, whereas this review centres on exercise and nutrition issues. Nevertheless, its well formulated structure was adopted to structure this review.

Search Methodology

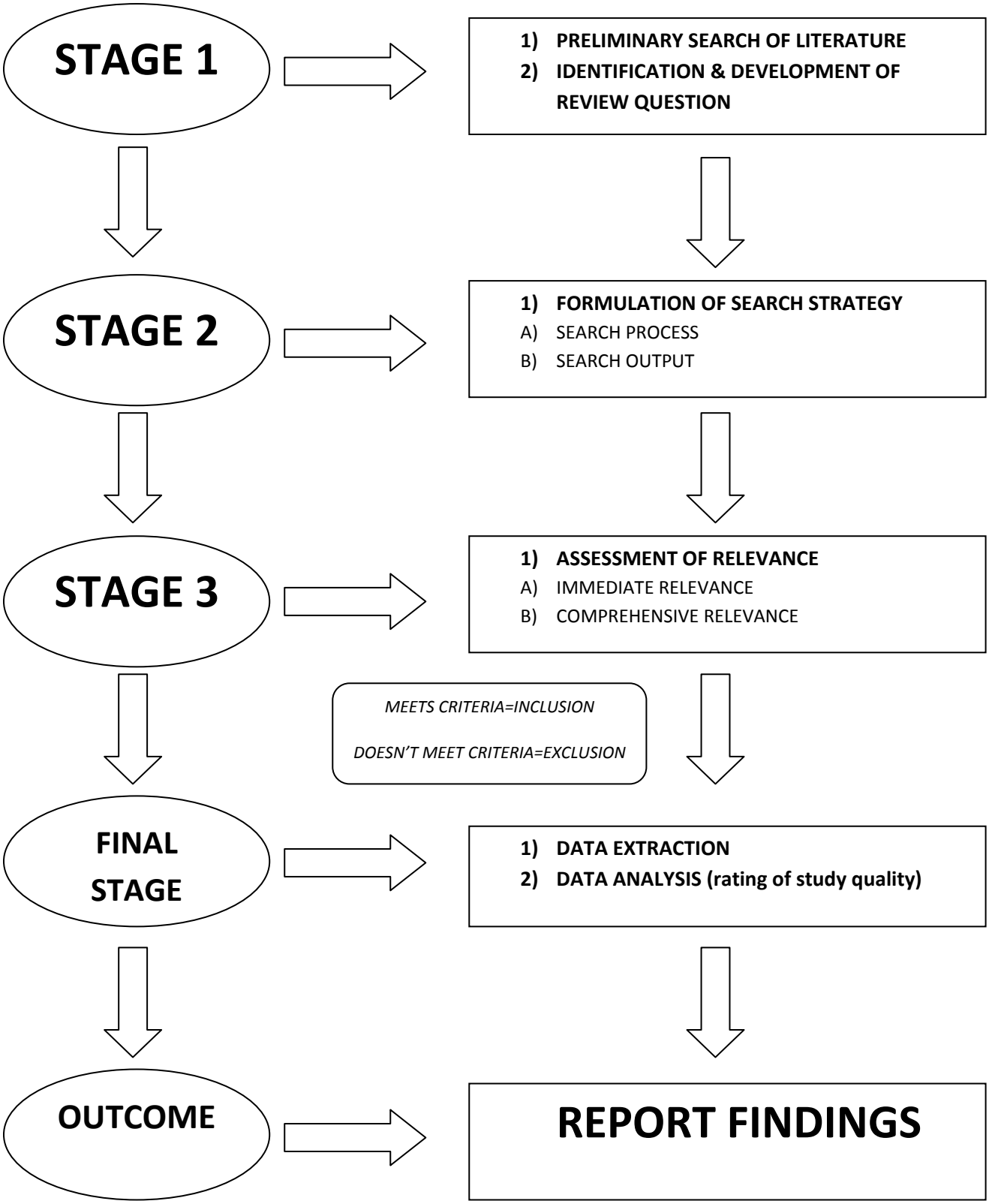

Figure 1. Summary of review structure

In order to undertake a systematic review, a structured approach was required involving key stages, (see figure 1). The first stage was a preliminary search of existing literature to obtain information on the topic, identification of an area within the topic that needs verification, and then development of the review question. The second stage was to formulate a search strategy involving search processes and search outputs of the literature. The third stage was to assess the references' relevance to the review question, by firstly assessing their immediate relevance with regards the initial criteria. Secondly, those references complying with the initial criteria underwent a second comprehensive relevance assessment, encompassing the inclusion criteria and methodological quality, once their respective full texts had been retrieved. The final stage involved data extraction and analysis of the included studies in the review, resulting in an overall rating score for study quality. The outcome of this allowed a final report to be formulated. Each stage will now be described in more detail.

STAGE 1

1) Preliminary Search of Literature

The initial approach to the systematic review was to perform a preliminary search of related academic literature to obtain information on the proposed topic, but also to establish the size and quality of the research evidence available. Google and Google Scholar as well as a small sample of relevant electronic databases, including Blackwell Synergy and PubMed, provided a broad and efficient filtered search and formed the basis for further specific searches.

The preliminary search indicated the orientation of the primary research in terms of its objectives and hence, provided an insight into its methodological designs. This information was useful in determining the direction of the review question. Furthermore, the search provided an opportunity to become familiar with search processes and resultant electronic databases. This familiarisation allowed for a more efficient search strategy.

2) Identification and Development of the Review Question

Protein requirements for individuals, particularly athletes engaging in sporting activity provided the original point of thought. There is a large amount of information available on this topic across many sporting disciplines accompanied by a high degree of controversy, in particular, when relating to protein's role as a performance enhancer. Maximising performance in all sports/disciplines is of paramount importance, so this is why protein's role in this instance must be addressed and possibly resolved. There is a strong possibility that protein requirements are increased above 0.8g/kg Bw/day with active individuals, such as strength athletes. This is due to no evidence suggesting otherwise as well as evidence supporting this claim, (Laritcheva *et al*., 1978; Burke *et al*., 2001; Hoffman *et al*., 2006). Therefore, it is important to identify the maximum requirements of these athletes in order to maximise their performance potential. Strength athletes were specified as the demographic under investigation, as they were perceived to train at high intensities, so are likely to maximise their protein intakes.

STAGE 2

1) **Formulation of Search Strategy**

A) Search Process

The search strategy was ultimately produced to provide a structured and efficient method of generating relevant research literature. As part of the search strategy, various sources of research literature were consulted. Specific search terms were required to initiate this process.

Relevant research was identified from three main sources; (i) electronic databases; (ii) 'grey' literature; (iii) reference chasing

Electronic Databases

Selection of Databases

Electronic databases were used as the primary source for research literature, as there is a vast amount of relevant information available from this source.

Six electronic databases were identified as appropriate through the university's electronic library of general resources for the Centre for Exercise and Nutrition Sciences. The reviewer interpreted the databases to embody a broad range of relevant research literature, e.g. in the fields of physiology, exercise science, sports medicine and sports nutrition. In order for a

thorough and comprehensive search to be performed, it was important to refer to each database independently, even though the search outcomes would carry a degree of replication between databases. Furthermore, referring to all six databases in the same fashion would both minimise the effects of database bias, (e.g. geographic bias) and enhance the search strategy reliability. Table 1 displays the consulted databases and their respective descriptions.

Table 1. Electronic Databases Explained

Database	Description
Blackwell Synergy	• One of world's largest journal publishers • Covers physical sciences, life sciences, medicine, social sciences, humanities
CINAHL Plus	• World's most comprehensive source of full text • Covers nursing, biomedicine, health sciences
IngentaConnect	• Global research gateway • Covers sciences, social sciences, humanities
Pubmed	• Service of U.S National Library of Medicine • Covers sports medicine, physiology, exercise science
Science Direct	• Contains over 25% of the world's science, technology and medicine full text and bibliographic information • Covers physical sciences, life sciences, health sciences, social sciences
Sports Discus	• Most comprehensive database of sport and fitness information • Covers exercise physiology, sports medicine

KEY:

No full text articles provided but wide selection of journals

Provides mainly full text articles but from a limited selection of journals

Selection of Search Terms

A list of keywords and phrases was compiled in order to search the databases. Firstly an initial search was performed using a sample of the databases, displayed in table 2, (Blackwell Synergy and Pubmed) using keywords deriving from the factors involved in the development of the review question. The keywords, 'protein', 'muscle' and 'strength' were used at this point, as they are key components of the review question; 'protein' being the nutrient in question; 'muscle' being the product of protein intake; and 'strength' being the athletes in question.

Initially, these key words were inputted into the databases separately to provide a broad output of references and to encompass all relevant references. However, there were many irrelevant references generated and an unmanageable number, (see table 2) so therefore, (a) more specific key words and phrases were required and (b) combinations of these words/phrases. Combinations have shown to increase sensitivity for sound studies allowing for more relevant studies to be retrieved, (Wilczynski & Haynes, 2005). The word 'protein' was the common denominator in these combinations, as protein is the most important factor in the research problem, (See Appendix 1: List of search terms). These additional words/phrases were identified when performing the preliminary search of literature, so to establish the type of terminology being used in this research area.

Table 2. Number of references generated from Initial search of sampled databases

Date of initial search-15th July 2008

		Blackwell Synergy	Pubmed
Initial search terms	**Protein**	268,720	4,103,310
	Muscle	71,246	694,578
	Strength	49,771	1,176,063

Searching

Each of the identified databases are preset to search for articles with these keywords/phrases and combinations based on their particular search field/index. These particular search fields/indexes are presented in table 3 together with their corresponding database. The specific key words/phrases were inputted into each of the database's preset search fields; however, as these search fields vary with each database, there would be a degree of inconsistency with the search strategy.

Table 3. Electronic databases and their respective search fields/indexes

Electronic database	Search field/index
Blackwell Synergy	Subject
CINAHL Plus	Keywords
IngentaConnect	Title, keywords, abstract
Pubmed	Topic
Science Direct	Title, abstract, keywords
Sport Discus	Keywords

As each database provides the option for combination searches, (i.e. advanced search) and therefore a more comprehensive search, combinations of these words/phrases were inputted. To further broaden this search, all search fields were included where possible. In the case of IngentaConnect and Science Direct, 'title, abstracts, keywords' were selected for

inclusion, as this search field provides the next best option. Also, with each keyword/phrase and subsequent combinations inputted, the number of search outputs were recorded together with any limitations or problems encountered during the search process. This was to provide prior information on any expected outcomes if this review process was to be replicated, (See Appendix 2 (i) for specific search strategy of each database).

In total 24 searches were performed with each database, (i.e. 13 keyword searches & 11 combination searches) and so a considerable number of outputs were produced. The number of outputs was noted but searches producing over 2,000 outputs were discarded, to reduce the number of irrelevant references and allow greater manageability for the review process. The outputs producing fewer than 2,000 references were saved for stage 3, (Assessment of Relevance). Figure 2 summarises the search process.

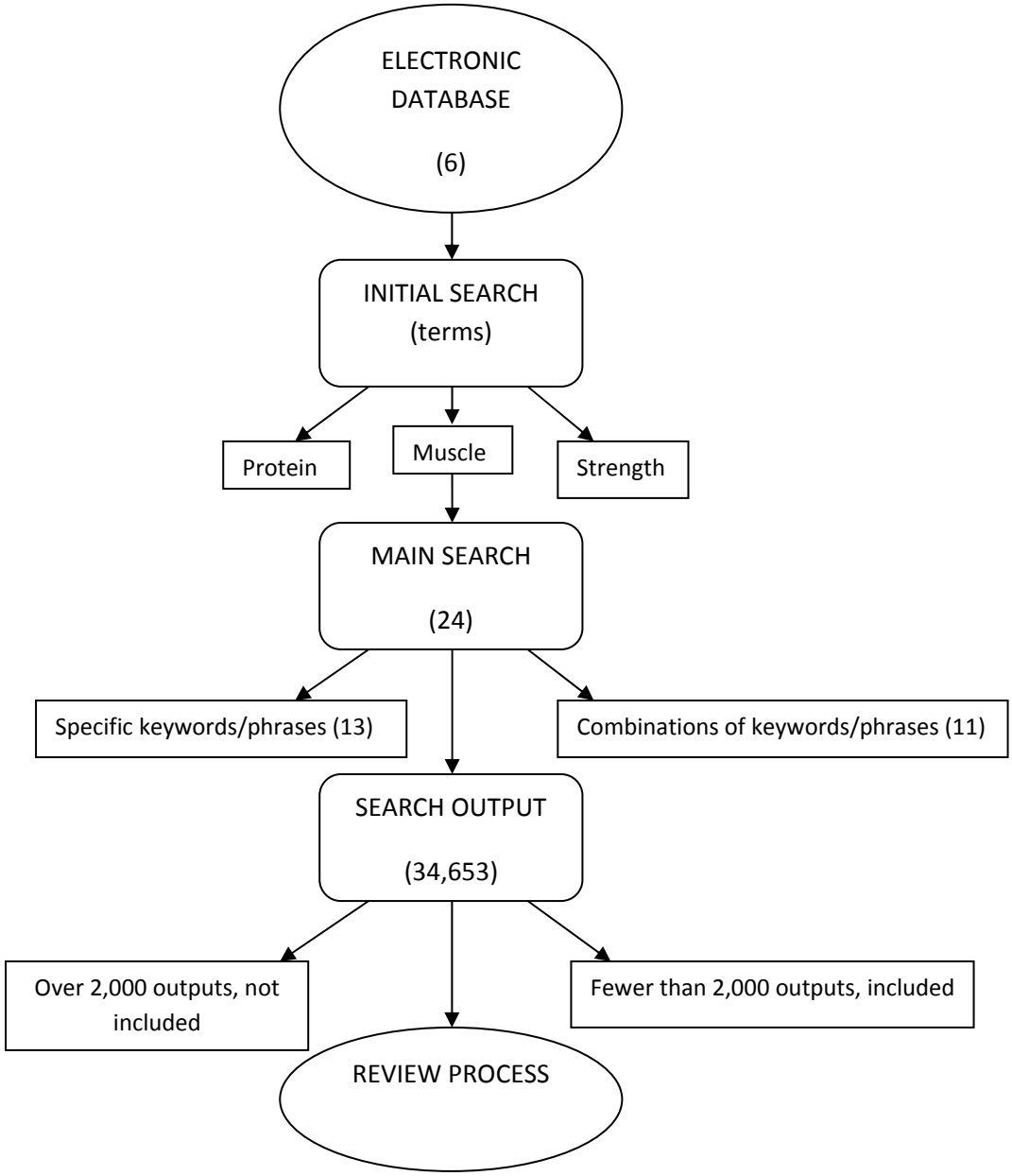

Figure 2. Flow chart summarising search process

Grey Literature

As well as searching electronic databases, 'grey' literature, (body of research that cannot be found through conventional academic channels) provides a source of references, such as theses, dissertations, newsletters and conference proceedings and abstracts. Also, as electronic databases also encompass grey literature, less emphasis was placed on this search strategy. Nevertheless, two sources were consulted, online search engines and hand searching.

Online Search Engine

The search engine, 'Scirus' was eventually chosen as the method employed to search for grey literature. This search engine provides different information types that specifically provide this source of research literature, and is likely to produce search outputs that cannot be generated by the electronic databases, (refer to table 4 for search strategy). The search engine, 'Google Scholar' was considered for use in tandem with Scirus, however, on its interface there was no option to search specifically for the various types of grey literature, so there were no accessible ways of retrieving relevant sources. As a result, this search engine was discounted from the search process.

Table 4. Characteristics of Scirus search engine and search strategy

Search interface	Description	Search strategy
Scirus	Science-specific search tool with results from over 200 million web pages, including sites that other search engines do not index.	24 searches in total using 13 keywords and 11 combinations, (See Appendix 1). Over 2000 outputs excluded from review process. Fewer than 2000 outputs included for review process. Keywords/phrases were inputted into the advanced search, 'all of the words' field, occurring in the title of the article. Combinations were also inputted into the advanced search, 'all of the words' field, occurring in title of article, of which the keywords, 'protein', 'muscle' and 'strength' should appear. (*All of the words were required to be in title of article to allow for a manageable number of outputs, as search engine is vast).* Information types selected were 'books', 'company homepages', 'conferences', 'patents', 'preprints', 'scientist homepages', 'theses and dissertations'. 'Articles' and 'abstracts' were not selected.

Hand searching

This method involves searching various forms of grey literature by hand for relevant information, although hand searching can also be achieved electronically, (i.e. online) by scanning for potential sources. The latter was implemented to specifically search the reference lists of the journals retrieved from the electronic database searches, (See 'reference chasing'). The other notable source referred to was the newsletter, 'Peak Performance', in particular, a special issue on sports nutrition, (issue-'Protein Matters!') as it closely relates to the review question. From this article the references were retrieved, (See Appendix 2 (ii) for number of search outputs of grey literature). When viewing the references, the reviewer's discretion was used in assessing their relevance, however, no less than two out of three of the initial search terms ('protein', 'muscle' and 'strength') were required in the reference title for further review.

Reference Chasing

Searches were performed on article reference lists when the comprehensive relevance assessment was completed, (after stage 3) to provide a manageable number of references. This search was also to increase the references' relevance to the review title and to ensure relevant references were not missed during the initial search process, (electronic database searches). As with the hand searching, reference titles containing no less than two out of three of the initial search terms ('protein', 'muscle' and 'strength') were included for the assessment of relevance. (See Appendix 2 (iii) for number of search outputs from reference chasing).

B) Search outputs

Collectively these methods yielded a total of 35,839 references, (See table 5). The search outputs as a result of reference chasing will be stated on completion of stage 3.

Table 5. Total number of search outputs from two sources

	Source		
	Electronic database	**Grey literature** Online search engine (Scirus)	**Grey literature** Hand searching
Number of search outputs	34,653	1,184	2
Total	35,839		

Date of search process- 15th July 2008

When searching each electronic database with the individual search terms, (13 keywords/phrases & 11 combinations) a total of 34,653 references were produced. The grey literature searches yielded 1184 references from the online search engine, 'Scirus', with two references deriving from hand searches.

It is evident that there were a vast number of references generated at this stage, most of which were irrelevant. Also, when searching across each of the six databases some studies generated were duplicated. This is likely to be due to the high number of search terms utilised.

STAGE 3

1) **Assessment of Relevance**

A) Immediate Relevance

In order to filter out relevant studies from the references generated from the search process, the 'immediate' relevance was assessed. This included developing relevance criteria that sought to specifically address the review question, without having to attain the full text of a study. References that adhered to this criterion would progress to the next tier of assessing relevance, the 'comprehensive' relevance. Table 6 encapsulates the criterions that were used to assess the immediate relevance. Criterion 1-3 was assessed by filtering the searches on the electronic databases. Criterion 4-5 was assessed by addressing the details of the abstracts.

The immediate relevance of the references was assessed in two parts. The first part involved filtering the search on the electronic databases, or in the case of hand searching, the reviewer retrieved the article and assessed it in accordance with the criteria. The second part involved manually assessing the abstracts of the studies meeting criteria 1-3 to determine their relevance to the review question; this part encompassed criteria 4 and 5. This criterion was also assessed using search filters where possible.

Table 6. Immediate relevance criteria

1	English-language study
2	Is a primary/empirical study
3	Research evidence published within the last 22 years, (1986-present)
4	Title and abstract ensues any relationship between protein and skeletal muscle of strength athletes, in particular; • Protein intake and muscle mass • Protein intake of strength athletes
5	Definition of keywords/phrases in abstract • 'Protein'-nutrient made up of amino acids consumed in habitual diet, as well as from supplements, (no exclusion bias). Encompasses total protein consumption. Therefore, all types of protein were included. • 'Muscle'- skeletal muscle or lean body mass or fat free mass used to create voluntary movement and in relation to resistance training. • 'Strength athletes'- individuals engaging in weight/resistance training activity to increase muscle. Includes bodybuilders, weightlifters, powerlifters.

Referring to table 6, studies had to be published in English, (1) and within the last 22 years, between 1986 and the present day, (3). This period of time was chosen, as the Australian Institute of Sport, (2006) states that extensive research has been acquired in the last 20 years on measurements of protein turnover and protein balance of strength athletes, at the time of their report. Studies conducted after the year of 2006 were included to update the potential findings of this review. Only primary/empirical research studies, (2) were deemed suitable for inclusion, as they produce data that is current and methodical. Notably, review studies were excluded. The title and abstracts of such primary research studies required a link in any form between all key aspects of the review question, (i.e. protein, muscle & strength athletes) in particular, the protein requirements of strength athletes with reference to muscle mass. Abstracts containing this link were considered relevant, (4). Titles and abstracts of studies not containing each of the following keywords were excluded from further review, 'protein', 'muscle' and 'strength athletes'. 'Strength athletes' synonyms including 'strength trained', 'bodybuilders' and 'weightlifters' were also used.

Assessing immediate relevance

The references generated from the electronic databases, (34,653) were assessed for studies meeting the criteria, 1-5. With regards the grey literature searches, references generated via Scirus followed the same filtering process as the electronic databases, although duplicated relevant references deriving from Scirus with these databases were excluded. The abstracts of the references produced through hand searching were retrieved at this point to assess criteria 1-5. Table 7 displays the total number of studies that met the immediate relevance criteria, (See Appendix 3 for the origin of the references included at this stage).

Table 7. Number of studies meeting the immediate relevance criteria

	Source		
	Electronic database	**Grey literature** Online search engine (Scirus)	**Grey literature** Hand searching
Number of search outputs	10	0	2
Total	12		

Note: references identified through reference chasing are not specified until after the comprehensive relevance stage, (completion of stage 3b).

.

Retrieval of full text of relevant studies

The 12 studies that met the immediate criteria underwent a 'comprehensive' relevance assessment. Firstly, the full text forms of these studies were required. These were retrieved from online sources, the university library and through inter-library loans connected to the British libraries document supply centre, (Boston Spa, West Yorkshire) when the study was not available locally.

B) Comprehensive Relevance

This section required the studies to meet more detailed criteria from three areas, the orientation of the study, (outcome measure) the methodological design, (interventions) and participant selection criteria. This is why the full text citations were necessary and to allow for an in depth assessment. Studies not meeting the criteria were excluded.

Orientation of study

For inclusion, each study had to show the effect of total protein intake on muscle, that is to say the quantity of protein and its effect on the size of muscle. Therefore, measurements regarding protein quantity needed to be evident in grams per kilogram of bodyweight per day, (g/kg Bw/day) exclusively, so to apply to all participants regardless of their bodyweight. This maintains external validity. More importantly this measurement was required in order to answer the review question, (i.e. outcome measure). Also, measurements regarding muscle size were required, for instance, measurements of length (cm) for cross-sectional area and/or weight (kg) for mass.

- Four articles were excluded on this basis;

Bamman, Hunter, Newton, Roney and Khaled, (1993) and Walberg *et al,* (1988) focus on the entire diet rather than protein separately. Van Zant, Conway and Seale, (2002) focuses on the effect of carbohydrate and fat rather than protein separately. Cribb and Hayes, (2006) focus on establishing the effect of supplement timing on skeletal muscle hypertrophy.

Methodological design

For inclusion, a comparison protein treatment group (s) was required in order to establish an effect.

- Four articles were excluded on this basis;

Cribb, Williams, Stathis, Carey and Hayes 2007) and Cribb, Williams and Hayes, (2007) did not include a comparison protein treatment group. Notably, creatine was used as a treatment group, which may produce confounding results due to its role as a muscle building supplement. Tang, Manolakos, Lysecki, Moore and Phillips (2007) and Hartman *et al*, (2007) use a carbohydrate treatment group.

For inclusion, total energy intake must be maintained across all treatments, (no significant change, $p>0.05$) to ensure the total protein quantity ingested is producing the effect and not changes in the energy intake.

An intense resistance training programme was required to be maintained during the study in order to maximise the protein requirements.

For inclusion, studies had to report a statistical significance equal to or lower than the alpha level of 0.05 for all results. This will determine whether the differences between the treatment groups are greater than can plausibly be attributed to chance.

Participant selection criteria

Only male participants were included, as more research evidence is available than on female participants.

For inclusion, the age range of participants was set between 18-35 years, as it is assumed that strength athletes are more active and likely to maximise their protein intake.

Components of a methodological design which are irrelevant in answering the review question

It should be noted that certain components of a methodological design were ignored, as they have no effect on the outcome of the review, so there is no issue in including or excluding these components, for instance;

- Control groups were not a necessary requirement, as participants act as their own controls when undertaking the different treatment groups, so that the effectiveness of each treatment group can be maximised. Additionally, no comparisons are being made between treatments, but rather the identification of a quantity.
- Crossover designs were not necessary as only strength athletes were included, so there were no issues with selection bias regarding treatment groups of which participants were randomly assigned.
- Sample size was not particularly important, so long as the treatment groups were randomly assigned, as the population demographic was pre-determined, thus, there is likely to be minimal variability between participants.

Outcome of comprehensive relevance

In combining the above assessments a total of eight studies were excluded, (See Appendix 4 for full references of excluded studies). This left a total of four studies that were relevant for the final stage of the review, (refer to table 8).

Table 8. Number of studies meeting the comprehensive relevance criteria

	Source	
	Electronic database	**Grey literature** Hand searching
Number of search outputs	2	2
Total	4	

Furthermore, the references of the 4 studies meeting the criteria were searched for more relevant studies that may have been missed through the entire search process. Relevant studies resulting from the initial search strategy underwent stage 3 (A) and (B) of the review process.

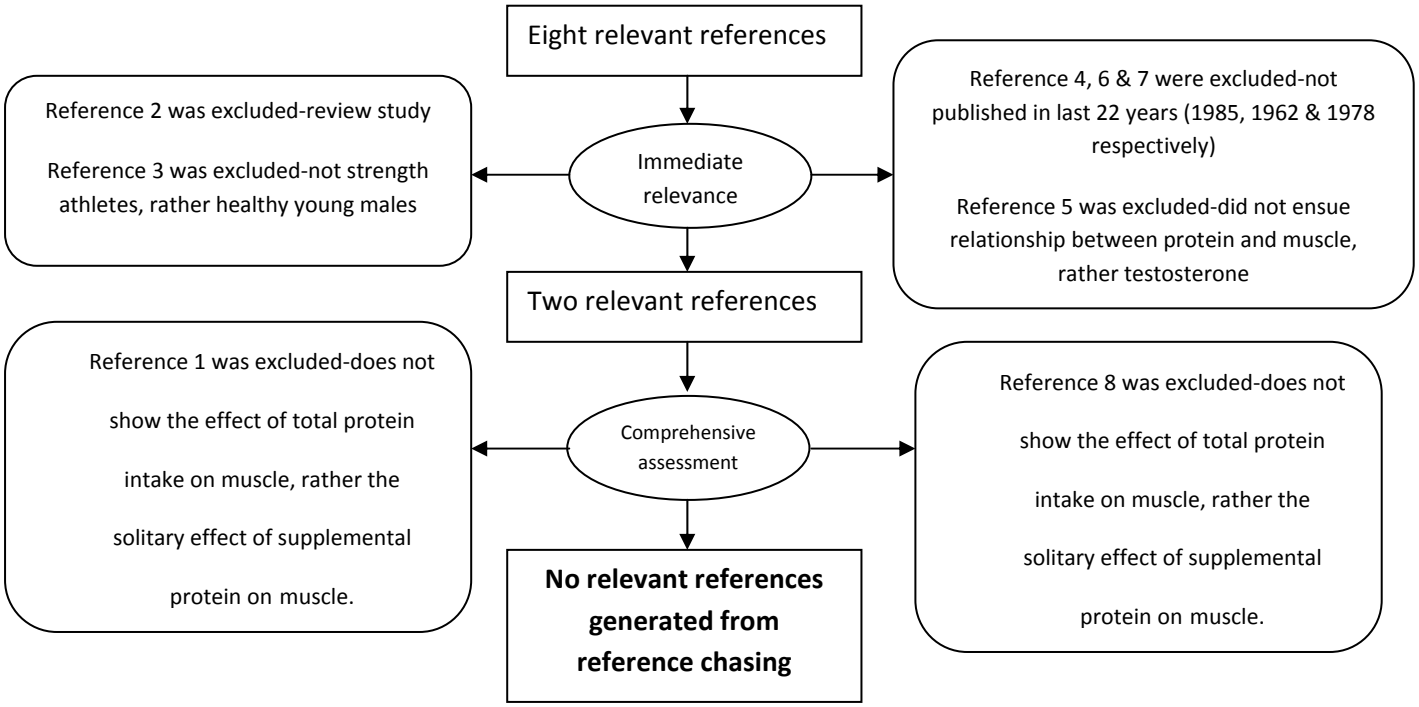

Figure 3. Flow chart summarising the process of filtering relevant studies from reference chasing

Final Stage

1) **Data Extraction**

Data extraction sheets were produced for the four studies extracted from the 'Assessment of relevance' stage, (See Appendix 5 for data extraction forms). The data extracted from the studies was standardised by being formatted into tables. Each study was concisely described in terms of its research design, sample, treatment groups, intervention, statistical analysis and outcome measure.

Results of included studies

A total of four studies involving 50 strength athletes with a mean age of 22.2 years, met the inclusion criteria. The group size of these studies ranged from 6 to 23 strength athletes, who were administered a certain quantity of total dietary protein each day and were assessed for any lean body mass changes, (outcome measure) ultimately determining the athletes' maximum protein requirements.

Outcome assessment

Muscle growth assessment was the most appropriate measure to assess protein requirements of these athletes, as the techniques involved make it much easier to detect experimental treatment effects, (Lemon, 1998). The most common assessment method to measure protein requirements involves measuring nitrogen balance, although it is not appropriate in assessing the maximum protein requirements, as it is used primarily for 'maintenance' measures rather than extreme measures. Furthermore, its validity as an assessment tool remains questionable, (Rand, Pellett & Young, 2003).

All four studies adopted highly reliable methods of lean body mass measurement, of which three studies utilise hydrostatic weighing, considered to be the gold standard of all densitometric methods, (Dempster & Aitkens, 1995). These studies include Lemon *et al*, (1992), Tarnopolsky *et al*, (1988) and Tarnopolsky *et al*, (1992). The remaining study (Hoffman *et al*., 2006) utilises body-dual energy x-ray absorptiometry, (DEXA) which is a highly precise method with a 2-3% margin for error, (Goffredsen, Jensen, Borg & Christiansen, 1986). These methods have determined that the maximum protein intake for all 50 strength athletes to be 1.4g/kg Bw/day, (Hoffman *et al*, 2006; Tarnopolsky *et al*., 1992) with the lowest maximum quantity being 1.05g/kg Bw/day, (Tarnopolsky *et al*, 1988) thus, the range between both extreme quantities was only 0.35g/kg Bw/day, which encompasses all four of the included studies.

Description and critical appraisal of included studies

Hoffman *et al*, (2006)

This study aimed to establish what effect protein intake had on body composition (lean body mass) of 23 collegiate strength/power athletes aged 18-24 years.

Subjects were assigned and counterbalanced to one of three separate treatment groups based upon their average weekly protein intakes;

Below recommended daily protein intake (BL), 1.0-1.4g/kg Bw/day (n=8),

Recommended daily protein intake (RL), 1.6-1.8g/kg Bw/day (n=7),

Above recommended daily protein intake (AL), >2.0g/kg Bw/day (n=8).

As no crossover design was present, a repeated measures design was essential. Though, no significant changes were observed in lean body mass as measured by DEXA post-test in any of the groups. So what can be deduced from the results is that 1.4g/kg Bw/day of protein intake is the maximum requirement, as no further muscle growth (measured in kg) occurred above this quantity;

BL= 76.8kg pre-test and 76.8kg post-test=no change,

RL= 73.9kg pre-test and 74.7kg post-test=0.77kg increase,

AL= 74.2kg pre-test and 75.3kg post-test=1.10kg increase.

Each treatment phase lasted 12 weeks encompassing a resistance training programme (4 days per week split routine), focusing primarily on developing strength/power. Also, caloric intake was similar across all treatment groups; BL= 3181 kcals; RL= 3127 kcals; AL= 3200 kcals.

Critical appraisal

The subjects in this study were elite strength athletes with at least two years of resistance training experience, so the results from this study would more closely reflect the objectives of the review question, and with a greater sample size, the statistical power of the results would be higher. However, it must be noted that this study is not randomised, in that the athletes were not randomly assigned treatment groups, so the validity of the results is somewhat questionable. Though, the results are still enlightening.

Athletes were categorised into one of three groups based on their weekly protein intakes, instead of being randomly assigned. A crossover design is essential in this instance since the athletes are already accustomed to their respective protein intake, thus, the likelihood of the results being invalid and unreliable is high. This is applicable even though the intervention occurs over 12 weeks, which provides a more than adequate time period for adaptation to the protein treatments.

The authors report that the caloric intakes of the athletes were low compared to those generally recommended for these types of athletes, thus, the ability of these athletes to make significant gains in lean tissue accruement is limited. Therefore, protein requirements may be unnecessarily increased.

Lemon *et al*, (1992)

This study investigated the protein requirements (intervention) of novice bodybuilders during intensive training, by assessing muscle mass changes by way of hydrostatic weighing (outcome measure).

In this randomised double-blind trial, increasing total dietary protein intake from 1.35g/kg Bw/day to 2.62g/kg Bw/day (treatment groups) did not significantly enhance muscle mass gains; lean body mass measures were similar between groups at 72kg, and 32cm for mid-arm circumferences and 50cm for mid-thigh circumferences.

All of the athletes undertaking the study (n=12) completed two-one month dietary treatment periods as part of a counterbalanced crossover design, separated by a seven day ad libitum diet washout period. This requires the random assignment of half the athletes to undertake one treatment group first before undertaking the other treatment group second. Each treatment phase was separated by a period of washout to prevent a carryover effect from occurring from the previous treatment phase.

Whilst undertaking the treatment phases over the two months, an intensive weight training programme was undertaken six days/week with the heaviest weight possible, and involving all major muscle groups. Importantly, the energy compositions of both treatment groups were similar on average; 4071 calories for the treatment groups using 1.35g/kg Bw/day of

protein intake and 4025 calories for the 2.62g/kg Bw/day of protein intake treatment group. Therefore, confounding dietary variables are avoided and it is the protein intake that is producing the effect observed with the outcome measurement.

A repeated measures analysis of variance was conducted to determine the effect of training and diet on changes in body muscle density (kg) and limb circumferences (cm) for both treatments. A statistical significance of $p \leq 0.05$ was set to establish a difference, but no significant difference was observed between the two treatment phases.

Critical appraisal

However, there are issues with the credibility of the study to produce valid and reliable results. The athletes included are defined as 'novice' bodybuilders in that the participants did not engage in a regular weight-training programme for 12 months before the study. This suggests their current training status is not as intense as more experienced strength athletes. This would be an issue when considering that protein requirements of an individual are related to intensity and volume of training, (Laritcheva *et al*., 1978; Burke *et al*., 2001; Hoffman *et al*., 2006) although on the contrary, training status is maximised with the use of heavy weightlifting targeting large muscle groups.

There is a limitation associated with the duration of each treatment phase, being that it was only one month in length, thus, athletes are not permitted sufficient time to adapt to the dietary protein intakes, and therefore maximise their protein requirements. The study's authors acknowledge this limitation but continue to state that a training effect occurred and thus, the outcome measure is valid. Consequently, 1.35g/kg Bw/day of protein may be too low as their protein requirements were not maximised.

Referring to the design of the study, the authors state that a limitation of the study was that it did not use a repeated measures design, possibly due to the time constraints of the study. However, the crossover design implemented is appropriate to this study, so that the effectiveness of each treatment phase can be determined within the same subject group. This is because they act as their own controls and the possibility of covariate imbalance can be nullified, (Senn, 1988) thus, enhancing internal validity. Although, the use of a 'counterbalanced' crossover design is inappropriate, as the order in which the different treatment phases are administered may have affected the outcome. Furthermore, the carryover between treatments is an issue, as the washout period was only seven days and therefore, a carryover effect may have affected the outcome as well.

A carbohydrate supplement comprising of maltodextrin was consumed with the lower protein amount, (1.35g/kg Bw/day) to make up for the energy deficit in quantity of the higher protein treatment group, (2.62g/kg Bw/day). The authors selected the carbohydrate supplement to compare with protein to enable sufficient fuel for intense exercise and to

contribute to an increase in lean body mass. However, the study's main objective was to determine the protein requirements of bodybuilders during training. The variable of carbohydrate may have confounded the outcome, as it may have reduced the protein requirements of the lower protein treatment group, (1.35g/kg Bw/day).

Tarnopolsky *et al*, (1988)

This study aimed to assess the influence of protein intake on lean body mass measured by way of hydrostatic weighing, on elite bodybuilders (n=6) aged 23-25 years, with at least three years of weight training experience.

The study did not randomly assign athletes to each treatment group, although the authors state the reason was to maximise dietary compliance, by starting on their habitual diets in the first experiment.

Experiment A involved using the athletes' normal diets based on a seven day food record collected before the initiation of the study, (2.77g/kg Bw/day). Experiment B was the altered protein diet consisting of 1.05g/kg Bw/day of dietary protein. However, this higher protein intake was not associated with an increase in body density (lean body mass) over the low protein intake, with body density being maintained at 1.08g/cm^3.

Thirteen days were assigned to each experimental phase with experiment B following immediately after the completion of experiment A. During the 26 days in total, the bodybuilders maintained their habitual exercise programme, but no further description of this exercise programme was given. Also, energy intake was maintained across both treatment phases in accordance with the bodybuilder's average habitual intake, (4800 kcals).

Finally, no significant difference was observed between the low and a high protein diet, as lean body mass was maintained.

Critical appraisal

Contrary to the previous study but as with Hoffman *et al*, (2006), this study recruited elite bodybuilders with more than three years weight training experience, so the results generated from this study are of more viable relevance, however, as the sample size is small (n=6) the statistical power of the results is reduced.

The treatment variables encompass high and low protein intakes, 2.77g/kg Bw/day and 1.05g/kg Bw/day respectively. However, the scope between the two protein diets in terms of their quantity was too high to establish a reliable outcome measure. A moderate protein diet would be an appropriate additional treatment variable to rectify this problem. Although, as the study found no significant increase in lean body mass between the low and high protein diets, no additional treatment group would be necessary.

The maintenance of muscle protein synthesis between the high and low protein intakes may be due to the athletes' recruited, as the bodybuilders are highly experienced and were in a period of 'maintenance'. Whereby, protein utilisation is more efficient on low protein diets, thus, the low protein treatment provided no further lean body mass accretion over the high protein diet.

As with the majority of studies, the duration of the interventions was short, only 13 days in this case. This may explain why lean body mass was maximised and maintained on such low protein diets, (1.05g/kg Bw/day).

Tarnopolsky *et al*, (1992)

This study's objective was to pinpoint the protein requirements for trained strength athletes, with lean body mass measures deriving from hydrostatic weighing.

In this randomised trial the athletes, (n=7) two rugby players, two football players and three weightlifters were assigned by counterbalance to three treatment groups;

Low protein intake (LP), (0.86g/kg Bw/day),

Moderate protein intake (MP), (1.40g/kg Bw/day)

High protein intake (HP), (2.4g/kg Bw/day).

Lean body mass or whole body protein synthesis was shown not to increase above the MP diet and in fact, decreased slightly from 77.7kg to 76.6 kg from the MP diet to the HP diet. Above the MP diet resulted in a nutrient overload, characterised by amino acid oxidation and no further muscle growth.

Each treatment phase lasted 13 days separated by two 8-day ad libitum washout diet periods. Further interventions involved circuit weight for the first ten days and whole body weight routines for the last three days. Energy composition of the diets administered during the treatment phases matched the athletes' habitual intake, determined from weighed food records. The LP diet (3595 kcals), MP diet, (3732 kcals) and HP diet (3723 kcals) were similar

in energy composition, thus, the treatment phase was more likely to be producing the effect. An alpha level of $p \leq 0.05$ was taken to indicate a significant difference, but no significant difference was observed from the MP diet to the HP diet.

Critical appraisal

As with Lemon *et al*, (1992) the timescale assigned to each treatment phase (13 days) was insufficient, so it may be that protein accretion was still occurring at higher protein intakes, thus, protein requirements cannot be maximised.

As highlighted previously with Lemon *et al*, (1992), a crossover design is appropriate in preventing covariate imbalance, although with three treatment phases there were two separate washout periods, which carries a greater risk of a carryover effect and thus, the outcome may not be valid. Therefore, this methodological design cannot be justified.

Varying levels of subject weight training experience from study to study will negatively affect the comparison of results. Subjects in this study only had 3-9 months of weight training experience. As bodybuilders with more weight training experience have shown to be more efficient at protein utilisation, (Lemon *et al*., 1992) the maximum protein intake in this study may be overestimated.

Regarding variables in this study, each diet consisted of 0-32% of energy intake deriving from whey protein. As this protein is of high quality relative to the dietary protein, protein requirements may have been underestimated, contrary to the previous point.

2) Data Analysis

Assessing the quality of included studies

The comprehensive relevance stage briefly examined the methodological design. However, a more in depth assessment of the quality of the methodologies was required for the studies extracted after the assessment of relevance, (stage 3). This involved assessing whether the studies used randomised controlled trials (RCT). RCT's are considered the most reliable from of research evidence, as spurious causality and bias are eliminated, that is to say that confounding factors that might affect the outcome are nullified, (Lachin, 1988).

'Quality' is considered a multidimensional concept encompassing the design, conduct and analysis of a trial, and with this, validity of findings generated by a study is an important dimension of this quality, (Juni, Altman & Egger, 2001). Validity can be subdivided into internal and external validity, (refer to table 9 for the components of internal and external validity) internal validity being the extent to which systematic error or bias is minimised in clinical trials, and external validity is the extent to which results of trials provide a correct basis for generalisation to other circumstances, (Campbell, 1957). When applying these terms to trials, internal validity implies that the differences observed between groups of subjects allocated to the different interventions may apart from random error, be attributed to the treatment under investigation. Whereas, external validity refers to external conditions, such as its application to other subject populations or treatment programmes, rather than solely the external conditions under investigation in one study, (Juni *et al*., 2001). It must be noted that internal validity is a prerequisite for external validity, so if the

results of a flawed study are invalid, then external validity becomes meaningless. Therefore, the assessment of internal validity was of paramount importance. This was achieved by using the most appropriate method of quality assessment.

Table 9. Summary of the components of internal and external validity associated with controlled clinical trials

Internal validity
• Selection bias: biased allocation to comparison groups • Performance bias: unequal provision of care apart from treatment under evaluation • Detection bias: biased assessment of outcome • Attrition bias: biased occurrence and handling of deviations from protocol and loss to follow up
External validity
• Patients, age, sex, severity of disease and risk factors, comorbidity • Treatment programmes: dosage, timing and route of administration, type of treatment within a class of treatments, accompanying treatments • Settings: level of care (primary to tertiary) and experience and specialisation of care provider • Modalities of outcomes: type or definition of outcomes and duration of follow up

(*Juni et al*., 2001)

Quality assessment methods

It is generally agreed that trial quality should be investigated in systematic reviews; however, there is no consensus on what methodology for quality assessment should be employed. As of yet, there is no gold standard answer to assessing the internal validity of a RCT. There are a large variety of scales available for measuring validity, but all have their own limitations. The most commonly used scales are described in more detail below.

Jadad Scale (Jadad, Moore & Carroll, 1996)

The Jadad scale uses a simple and easy approach that incorporates the most important elements of methodological quality. It is a widely used method of assessment for clinical trials, as it is a validated measure of quality, and time to completion of the assessments is only 10 minutes. It utilises a five point scale in which points are awarded if the study is described as;

- Randomised (+1)
- The means of carrying out randomisation is described and appropriate (+1)
- The study is described as double-blinded (+1)
- The means of double-blinding is described and appropriate (+1)
- There is a description of withdrawals giving the number and reasons for withdrawals (+1)

However, points are deducted if;

- The method to generate the sequence of randomisation is described and is inappropriate (-1)
- If the method of double-blinding is described and is inappropriate (-1)

Juni *et al*, (2001) reinforce the inclusion of randomisation, blinding and withdrawals, (attrition) in the analysis by stating it should always be assessed when evaluating the quality of any RCT, (Schulz, 1995). Also, as allocation concealment ensures proper randomisation,

an inadequate concealment of allocating treatment groups and a lack of double-blinding would result in an exaggeration of treatment effects, (Schulz, Chalmers, Hayes & Altman, 1995). Therefore, the Jadad scale consists of components that are essential in potentially validating the answer to the review question.

The creators of the scale achieved an interrater coefficient of agreement of 0.66 with a 95% confidence interval rating of 0.53 to 0.79 for the whole scale. These ratings indicate that this scale is a highly reliable method of assessing trials. The interrater agreement of Jadad's scale is consolidated by a study by Clark, Castro, Filho and Djubelgovic, (2001) who report similar agreements for the whole scale, between 0.69 and 0.81.

PEDro Scale

The PEDro scale is a validated evaluation instrument initially developed to rate the quality of RCT's on PEDro, the Physiotherapy Evidence Database. As with the Jadad scale, the PEDro scale is a checklist that examines the internal validity of trial quality although this scale also assesses its 'interpretability' or statistical reporting. The scale uses an 11-item checklist which yields a maximum score of 10 points if all criteria are satisfied. Criteria 1-8 assesses the internal validity specifically within the study design, such as random allocation, concealment of allocation, comparability of groups at baseline, blinding of participants, administrators and assessors, adequacy of follow up and analysis by intention to treat. Criteria 9-10 examine between-group statistical comparisons and descriptions of both point estimates and measures of variability.

As with the Jadad scale, the PEDro scale has tested reliability data with an interrater coefficient of agreement of 0.68. This is a similar score to those reported by other commonly used quality scales, in particular, the Jadad and Chalmers' scales, (Maher, Sherrington, Herbert, Moseley & Elkins, 2003).

A study by Bhogal, Teasell, Foley and Speechley, (2005) states that the PEDro scale provides a more comprehensive measure of methodological quality than the Jadad scale, but only when double-blinded studies are not possible.

Chalmers Scale (Chalmers, Smith &Blackburn, 1981)

The Chalmers scale was one of the first quality assessment scales developed. It comprises of four subscales where only three are integral components of the overall score. These integral subscales include methodological aspects, statistical analyses and presentation of results. The first subscale includes the author and journal, names, year of publication, funding source, country and affiliations masked. The scale includes 27 items in total over the four subscales, ranging in scores from 0 to 100. Fourteen items or 60% of the scores are designed to assess both internal and external validity. Contrary to the previous two scales, the Chalmers scale is not specifically designed for validity assessment, although it does deal with internal and external validity as well as statistical evaluation and validity of the presentation results. However, because of its lack of validity assessment and with its subscale validity still questionable, it has mostly been used for assessing study eligibility, (Berard, Andreu, Tetrault, Niyonsenga & Myhal, 2000).

The Chalmers scale achieved an interrater coefificent of agreement of 0.66 with a confidence interval rating of 0.55 to 0.79. This scale is also a highly reliable method of assessing trials with similar levels of agreement to the Jadad and PEDro scales. Therefore, the Chalmers scale should be considered for assessing the trials used in this review; however, its weighting assigned to internal validity may be insufficient.

CONSORT Scale (Begg et al., 1996)

The CONSORT statement was developed in the mid nineties by an international group of clinical trialists, statisticians, epidemiologists and biomedical editors to improve the quality of reports of RCT's. The statement lists 21 items that should be included in these reports including the description of the study format, description of protocol, method of randomisation, blinding and follow up, and appropriate description of the analysis and discussion of results. In essence, the CONSORT statement does not score randomised trials according to specific criteria like the previous three scales described above, but instead reports key information within randomised trials that relate to the study quality.

This method of quality assessment relies more on the interpretation by the reader with regards its reliability, rather than a score defining its quality. A statement rather than a score means that no important information is omitted. Also, the authors of these trials often hide their procedures behind the word, 'randomised' although this method of assessment requires details of the randomisation procedure to be present.

Downs & Black checklist (Downs & Black, 1998)

The Downs and Black checklist was developed to assess both randomised and non-randomised studies. It encompasses internal validity (bias and confounding) and power as well as external validity, and alerts the reviewer to any strengths and weaknesses of a trial. The checklist consists of 26 items spread across five sub-scales; reporting, external validity, internal validity (bias), internal validity (selection bias) and power.

The majority of answers required 'YES' or 'NO' scoring 1 or 0 respectively. Although, one item on the reporting sub-scale scores 0 to 2 and the item on 'power' was scored 0 to 5. Therefore, a maximum score of 31 was assigned to this checklist. A score of 23 or over (75%) is an acceptable threshold to indicate a high quality study.

The creators of the scale achieved a good inter-rater reliability score of 0.75 for both randomised and non-randomised studies, a high internal consistency of 0.72 and a high retest reliability of 0.88. These results show that this checklist performs highly even after revision, and can be proclaimed a highly reliable method of assessing these forms of trials on a large majority of research interventions.

Review of methods of assessment

The five methods of assessment described above, (Jadad scale, PEDro scale, Chalmers scale, CONSORT statement & Downs and Black checklist) are commonly used in systematic reviews as they have been found to be among the most valid and reliable assessment tools for RCT's, (Olivo *et al*., 2008).

Jadad Scale

The main justification for the selection of the Jadad scale is that it includes properties that are vital in evaluating the quality of any RCT, (i.e. randomisation, blinding and handling of subject attrition), (Juni *et al*., 2001). As well as incorporating the most important properties of methodological quality, it is simple and easy to administer which in essence limits the potential for evaluating bias by the reviewer. Finally, the scale presents the best validity evidence and has been tested for reliability in different settings, thus, it can be viably applied to a wide range of areas of review. Furthermore, unlike any other scale it is the only one that was constructed according to psychometric principles, (considers all aspects of validity), (Juni *et al*., 2001).

However, the Jadad scale does have limitations. The reviewer is left to interpret whether the methods of randomisation and blinding were appropriate, as the guidelines are too lax, thus, influencing the quality rating score of a RCT, (Bhogul *et al*., 2005). Also, the scale lacks an element of allocation concealment which can have severe negative effects on internal validity, regardless of whether provisions are made for randomisation and blinding.

PEDro Scale

Where the Jadad scale proposes a 3-item approach, the PEDro scale utilises an 11-item scale including both internal and external validity measures. This can provide a more comprehensive assessment of methodological quality, which is further exemplified by the fact that the guidelines are not open to interpretation, rather 'yes' or 'no' answers. As with the retest reliability of the PEDro scale proved to be fair to good, (Maher *et al*., 2003) this scale is versatile in its application.

The appropriateness of this scale to this review is questionable as it focuses on aspects of a study's methodological quality, such as statistical reporting besides internal validity. As internal validity assessment is of upmost importance to this review, a scale solely focusing on assessing this entity would be more desirable.

Chalmers Scale

The Chalmers scale is among the most reliable methods of quality assessment along with the Jadad and PEDro scales, with an overall interrater agreement coefficient of 0.66. Even though the scale contains the subscale measuring protocol and methods, which are the most important aspects to consider when evaluating study validity, (Berard *et al*., 2000) there are many flaws associated with it.

The time length to complete the quality evaluation of the Chalmers scale is approximately 40 minutes on average, whereas, the Jadad scale takes only 5 minutes. With such a reduced timescale to completion, the potential for evaluation bias is reduced significantly, but is likely to be evident with this scale. Also, the Chalmers scale is not specifically designed for validity assessment as the validity of the subscales remains to be determined, (high variability between the reliability estimates of the subscales). Therefore, this scale may not be the most reliable when applied to the context of this review. The scale mainly assesses study eligibility as it gives an overall quality score. However, even a careful assessment of the scale with studies scoring high, the Chalmers scale can still mask the presence of low internal validity (systematic bias) that may exist in the RCT's, (Berard *et al*., 2000). This is not a desirable quality of an assessment scale as establishing the level of internal validity is essential in assessing the quality of a RCT. The Chalmers scale is therefore not an appropriate method of assessment with other scales proving more appropriate, notably the Jadad scale.

CONSORT Statement

The CONSORT statement lists 21 items that should be stated in a RCT. If any of these items are missing, the RCT's validity is likely to be reduced. As the CONSORT statement is a report, it is apparent to reader the exact properties of the RCT in question. This allows the reader to judge whether the findings are likely to be reliable, however, this is open to interpretation and may be affected by the level of understanding of the reader. Therefore, systematic error (bias) may arise, what makes a reliable RCT and what does not? Alternatively, such a report means inadequacies of a RCT cannot be hidden by omitting important information, whereas with the scales the assessment results can score with no explanation. Therefore, the scales are likely to ignore important information, and so the CONSORT statement may provide a more comprehensive assessment.

However, the checklist engulfing the CONSORT statement only applies to the most common design of RCT's. Modification of the checklist would be required for other types of trials, such as crossover trials and those with more than two treatment groups. Therefore, this method of assessment lacks versatility and so will not provide an appropriate measure of all RCT's. As well as being inflexible the items on the checklist would benefit from greater explanation, thus, systematic error may arise from the potential misinterpretation of what the items ensue. As with the PEDro and Chalmers scales', the lack of appropriateness associated with the CONSORT statement in assessing RCT's, may prove its downfall for selection in this review.

Downs & Black checklist

The main justification for the selection of the Downs and Black checklist is that it can judge the methodological quality of both randomised and non-randomised studies. This is particularly relevant as half of the studies extracted for the review were non-randomised but relevant, so their inclusion would be desirable. Furthermore, the validity and reliability of the checklist scores highly. This is typified by the fact that its performance as an assessment tool was as good as other established checklists. Additionally, little difference is observed in performance with randomised and non-randomised studies.

Unlike many other scales/checklists used to assess methodological quality, this checklist provides a structured profile of a research paper. This alerts the reviewer to any methodological strengths and weaknesses and so allows the reviewer to interpret whether or not the checklist is appropriate to the review. This to some extent is addressed as part of the CONSORT checklist with descriptions of study format and protocol, although the Downs and Black checklist addresses them to a greater extent and is therefore more comprehensive.

However, the checklist does carry limitations. A lack of sufficient definitions of some items causes difficulties in interpretation and therefore scoring, e.g. 'are the interventions of interest described?' In this case what kind of interventions?

With regards the external validity, the assessment of this sub-scale presented poor reliability possibly due to its lack of items, only 3. Also, as mentioned in the previous point, the lack of conciseness in the items may have contributed to poor reliability, as interpreting these items have proved difficult.

Ultimately, this assessment tool is newly developed so there is still a lack of understanding as to the impact of each dimension (sub-scale). This may be the defining factor in whether this checklist can be encouraged for routine use.

Selection of method of quality assessment

The method of quality assessment selected for assessing the studies meeting the comprehensive relevance criteria was the 'Downs and Black checklist', (See Appendix 6 for description and guidance for use). This scale was selected on the premise that it is the only measure of quality assessment that assesses non-randomised studies as well as randomised studies. Also, other methods of assessment have proved inappropriate so on this basis they were excluded. Furthermore, the Downs and Black checklist provides a more comprehensive assessment of the methodology, which encompasses the internal validity. 13 of the 27 items are assigned to internal validity (bias and selection bias).

The Jadad scale was strongly considered as the assessment tool for this review, as it is the only validated measure of quality. Also, it centres entirely on the internal validity of a trial, an important component that is necessary in any trial. However, the Jadad scale cannot be applied to measuring non-randomised trials. The scale is also irrelevant to the review topic as it primarily assesses clinical trials of the general population, whereas the trials within this review use subjects of a specific demographic. The PEDro, Chalmers and CONSORT assessment tools do not assign enough emphasis to measures of internal validity, so were excluded from selection on this basis.

The design of any trial aims to test to see if an association exists between the intervention and outcome measure and to minimise flaws in the design, so that no bias arises when measuring this association. The weighting assigned by the checklist to measure the quality

of the design is relatively high with 10 of the 27 items, (Reporting) so bias is likely to be reduced.

Even though there are issues with interpreting the items within this assessment tool, 'YES' and 'NO' answers allow an efficient system of measuring, thus, limiting the potential for evaluation bias.

Ultimately, this assessment tool provides the opportunity to assess the methodological quality of non-randomised studies, which would have not been possible prior to the creation of this checklist. Moreover, in many areas of healthcare, the few randomised trials that exist have been poorly executed, (Downs & Black, 1998). This checklist is the most appropriate quality assessment method for assessing such trials, as it acknowledges and assesses all aspects of a study's design.

Quality Assessment

The quality assessment using the Downs and Black checklist, (See Appendix 7 for quality assessment scores of studies) shows that 3 out of the 4 studies can be classified as high quality, (75% or 23/31). These studies are Hoffman *et al*, (2006), (24/31), Lemon *et al*, (1992), (24/31) and Tarnopolsky *et al*, (1992), (25/31). The remaining study which is non-randomised (Tarnopolsky *et al*, 1988) scored 20/31, so was excluded from the review process. The other non-randomised study by Hoffman *et al*, (2006) is of high quality. This is contrary to popular belief that non-randomised studies are difficult to assess for methodological quality, as their research evidence is deemed invalid and unreliable.

When examining the individual subscale scores for each of the three included studies, it appears that the scores for the 'reporting' subscale are high across all studies, with 8/11 being the lowest score for both Hoffman *et al*, (2006) and Lemon *et al*, (1992); Tarnopolsky *et al*, (1992) scored 9/11. This suggests the designs of the studies are of high standard and that sufficient information is available with regards the intervention, so there is less likelihood of bias from occurring.

Bias was also examined when assessing the internal validity of the studies. Internal validity is an essential dimension of quality, (Juni *et al*, 2001) thus, it is important that this subscale scores highly. Both Lemon *et al*, (1992) and Tarnopolsky *et al*, (1992) scored highly with 12/13 and 10/13 respectively, but the score achieved by Hoffman *et al*, (2006) was not as high. This is due to the study being non-randomised but also, these types of studies do not

blind subjects or investigators from receiving and measuring the intervention, respectively, so bias may have occurred in this instance.

Overall, all studies have shown to have a high level of internal validity, as the likelihood of bias has been reduced and so they can be referred to as high quality research evidence. Therefore, their outcome measures are likely to be valid and reliable and thus, the review question can be answered sufficiently.

OUTCOME

Reporting of findings

The systematic searches of electronic databases and grey literature resulted in the location of three studies, two randomised and one non-randomised, on the maximum protein requirements of strength athletes. These studies are of high quality as assessed according to the criteria of the Downs and Black checklist. Yet, it must be noted that the protein requirements of strength athletes has an extremely sparse evidence base, so there is no convincing evidence supporting the maximum protein requirements of these athletes.

From the studies accepted into the review process, it is deduced that the maximum protein requirements of strength athletes is 1.4g/kg Bw/day, as this is the highest quantity of all three studies. Interestingly, all studies provide a similar quantity. Hoffman *et al*, (2006) shows the effect of protein intake on body composition in strength and power athletes, and found that 1.4g/kg Bw/day of total protein intake provided the maximum threshold above which no further significant muscle growth was evident. This quantity was matched (1.4g/kg Bw/day) by a study evaluating the protein requirements of trained strength athletes, (Tarnopolsky *et al*, 1992). Lemon *et al*, (1992) investigated the protein requirements of bodybuilders by measuring any muscle mass changes. They found that no significant muscle growth occurred above 1.35g/kg Bw/day of total protein intake.

These quantities equate to a moderate intake which is contrary to what protein requirements are believed to be essential for maximising muscle mass in strength athletes, high protein diets greater than 2.0g/kg Bw/day, (Tarnopolsky, MacDougall & Atkinson, 1988; Steen, 1991). This amount is considered excessive, as protein not being used for muscle protein turnover is simply oxidised, (Lemon, 1998).

This systematic review differs from other notable review papers. Tipton and Wolfe, (2004) and Lemon, (1998) found the maximum threshold to be 1.7g/kg Bw/day of total protein intake, (moderate to high) yet there is no indication as to how they came to arrive at this conclusion, as these reviews are not systematic and so show no search strategy.

Such a low maximum requirement for strength athletes deriving from this review, may simply be explained by the interpretation that exercise training increases the efficiency of protein utilisation, thus, making increased protein intake unnecessary, (Butterfield & Calloway, 1984). However, this study used endurance exercise as its intervention, so it is yet to be determined whether the same effect will occur with strength training, although the same theory is likely to apply. Lemon, (1998) addresses this point and suggests that more dietary protein is required to build muscle than to maintain it. This seems feasible since strength athletes would have trained over a number of years to maximise their muscle building potential and during this phase, their protein intake may have indeed been much higher than even 2.0g/kg Bw/day. It is therefore possible that at the time of the intervention in the studies, the strength athletes were in a ‘maintenance’ phase. Although, longitudinal

studies are required to determine whether such chronic strength training reduces the increased protein need required during the initiation phase of training, (Lemon, 1998).

What do the results mean?

Overall, the results of the three studies included in this review are positive when reviewing the rationale and associated relevance of the review.

Firstly, when referring to 'rationale 1', it is deduced that higher protein requirements would be necessary to train at higher intensities, so to maximise these athletes' physical performance. The results of this review do not support this rationale when relating to a greater muscle mass and its role in maximising performance, as excess protein (above 1.4g/kg Bw/day) is simply oxidised.

'Rationale 2' relates to the risk of developing health problems with high protein diets. As high protein diets are not necessary for these athletes, the risk of them developing health problems should be nullified.

'Rationale 3' relates to how these athletes' require protein in convenient forms (drinks and snacks) on top of their habitual protein diets to meet their protein requirements. This is due to their hectic lifestyles or frequency of training. The results of this review do support the consumption of supplements but to a lesser extent than previously thought. Certainly,

protein supplements have their place in hectic lifestyles, although moderate protein requirements can be achieved easily through habitual protein diets of strength athletes.

‘Rationale 4’ relates to acquiring protein supplements and their associated financial burden, particularly if the strength athletes are amateur, and so may not have the necessary disposable income available to purchase them. This rationale closely relates to ‘rationale 3’ in that it questions the need for supplements. As supplements are not a necessary requirement there would be no financial burden, however, as previously stated a moderate protein requirement may require the acquisition of supplements but only with busy lifestyles. Though, this may not relate to elite strength athletes who do not work long hours in addition to training.

Limitations of included studies

As quality of reporting as assessed by the Downs and Black checklist was generally good for all studies, there were confounding factors and limitations in the design between trials that may negatively affect the close proximity of all three studies. Therefore, a convincing answer to the review question is difficult to attain.

The total energy intake differs between studies; Hoffman *et al*, (2006) allocates 3100-3200 calories to energy intake, Lemon *et al*, (1992) allocates 4000-4100 calories and Tarnopolsky *et al*, (1992) allocates 3600-3700 calories. Even though the range between studies is only 1000 calories, a higher energy intake can have a protein sparing effect but still stimulate muscle development, (Lemon *et al*, 1992). Therefore, studies such as Lemon *et al*, (1992) with a higher energy intake cannot be made comparable to Hoffman *et al*, (2006) with a lower energy intake, even though the results are similar.

Another factor that potentially confounds the results between the studies is the athletes' body mass. Even though the outcome measure accounts for differences in body mass as it is a relative measure, the amount of protein required solely for the purposes of lean body mass cannot be accounted for, so the results between studies may still vary. Body mass can also be an issue within trials. Changes in body mass during the studies due to muscle growth or loss can affect the result post-test, as protein requirements will inevitably change.

The length of time for the intervention varies considerably between studies; 12 weeks are assigned by Hoffman *et al*, (2006), 4 weeks by Lemon *et al*, (1992) and only 13 days by Tarnopolsky *et al*, (1992). Therefore, whether sufficient time was allocated in order to adapt to the treatments may be an issue with the study by Tarnopolsky *et al*, (1992). Indeed, the protein requirements of these athletes' may prove higher with a sufficient adaptation period.

The experience levels of the athletes' between studies may affect the results collectively. From the athletes' in Hoffman *et al*, (2006) with at least 2 years of resistance training experience to the athletes' in Tarnopolsky *et al*, (1992) with only 3-9 months, the effect of the treatments may be less obvious with the more experienced strength athletes. This is because most of the potential muscle gain would have already occurred. However, this appears to be not the case as results are similar across all studies, yet the lack of an adaptation period in Tarnopolsky *et al*, (1992) may still invalidate this point.

Having considered the effect of these limitations, it appears the main discrepancy between the studies is the length of time each study designated to the intervention. To resolve this limitation, a longitudinal endpoint study might be the best option. However, strict control of all aspects of an athletes' life (e.g. diet, training, rest, travel) would be necessary to obtain more reliable results, but this may prove virtually impossible to conduct.

IMPLICATIONS FOR FUTURE RESEARCH

All three trials do demonstrate that moderate protein intakes represent the maximum protein requirements of a strength athlete. However, longitudinal studies should be at the forefront of any further research to consolidate these findings. Although, care should be taken in controlling for all aspects of an athlete's life, in order to minimise confounding factors because they cannot be completely eradicated.

As between trial differences are prominent as highlighted previously, (See Background- Problems associated with identifying maximum protein requirements) and those observed in this review, such as body mass, the level of experience of the athlete, and energy intake, a true comparison and decisive answer cannot be achieved. These components of a design cannot be sufficiently controlled for, so any future research on this topic can only be at best a guideline value.

With regards measuring the maximum protein requirements, there is no preferred standardised measure. Therefore, there is a further problem with generalising these results to strength athletes. This problem lies in the fact that there are various methods of measuring lean body mass and these methods differ between trials, (e.g. hydrostatic weighing, circumference measures and CAT scans). So, each method may generate a different outcome measure and thus, replication of findings will be made difficult. This problem needs to be resolved in order to produce more reliable results.

Given the variations in study designs on top of the lack of research evidence from this area, there is no indication that the findings are reliable or indeed will be in the near future.

STRENGTH AND LIMITATIONS OF THIS REVIEW

Designing a search strategy to locate research evidence of protein requirements in strength athletes was difficult because of the large number of potential descriptors for such studies, and this is a notable weakness of this systematic review. Although, this study used 24 different search descriptors so the difficulty in locating these studies was minimised. This was even though some studies were not adequately indexed by the databases that were searched. These databases include Blackwell Synergy, IngentaConnect and Science Direct, which did not produce any relevant studies. If the review was to be replicated, the electronic databases selected should primarily encompass sports and exercise science and sports nutrition, which were not covered sufficiently in the previously stated databases. But, overall the main searches together with the preliminary searches appear to result in no more hidden studies related to the topic of the review.

A notable strength of the review was the thorough search strategy implemented, particularly with regards using a two tier relevance system, to provide a more comprehensive assessment and filter out the most relevant studies. However, a limitation of the search strategy was that a large number of studies were discarded, as searches according to each search descriptor producing over 2000 results were ignored, and so no further processes were undertaken with these results. This was to prevent the filtering process from being too time consuming.

Another associated limitation was the small sample of search outputs meeting the immediate criteria. Even though only 12 references met the immediate relevance criteria from a possible 35,839, the keywords and phrases required in the title of the references were considered general yet relevant enough to meet the aims of the review, and to potentially encompass all of the relevant research literature, so no alternative search strategy was necessary.

In an attempt to counterbalance the potential loss of relevant studies, the reviewer introduced the method of 'reference chasing,' which proved to be a strength of the review process. This method was applied to those studies meeting the criteria for inclusion of both tiers of relevance. This method not only provided further studies for review that were not otherwise identified through database searches, but these studies proved to be highly relevant. As a result, one of the three studies included for review was identified through this method, though the remaining two studies were also identified this way, as well as being initially generated through database searches.

Overall, this systematic review was comprehensive, in that it successfully identified studies that attempted to answer the review question.

CONCLUSION

The maximum protein requirement of strength athletes, above which no further contribution to lean body mass would be evident, is 1.4g/kg Bw/day, which equates to a moderate protein intake. This quantity is contrary to the popular belief amongst these athletes that high protein diets are necessary, (over 2.0g/kg Bw/day), (Tarnopolsky, MacDougall & Atkinson, 1988; Steen, 1991). Acknowledging this quantity will not only increase physical performance, but will have the added benefit of reducing the risk of health problems associated with high protein diets and decrease the financial burden of acquiring protein supplements.

The reliability of this result remains questionable due to two defining factors. Firstly, there is an extremely sparse evidence base supporting this quantity in strength athletes. Secondly, studies with similar objectives have dissimilar methodological designs, which encompasses total energy intake, body mass, experience levels of athletes, length of time of intervention and method of measuring the outcome. In combination these factors significantly affect the integrity of the result.

Nevertheless, this quantity can be used as a guideline value for strength athletes engaging in intense resistance training, but has yet to be determined as a reliable maximum requirement.

References

Allen, L.H., Oddoye, E.A. and Margen, S. (1979). Protein-induced hypercalciuria: A longer term study. *American Journal of Clinical Nutrition*, 32, 741-749.

American Dietetic Association, Dieticians of Canada and the American College of Sports Medicine. (2000). Position stand: Nutrition and athletic performance, *Journal of Medicine and Science in Sport and Exercise*, 32, 2130-2145.

Astrand, P.O. and Rodahl, K. (1977). *Textbook of Work Physiology*. New York: McGraw-Hill.

Australian Institute of Sport. (2006). *Protein-are you getting enough?* Department of Sports Nutrition.

Begg, C.B., Cho, M.K., Eastwood, S., Horton, R., Moher, D., Olkin, I., Rennie, D., Schulz, K.F., Simel, D.L. and Stroup, D.F. (1996). Improving the quality of reporting of randomised controlled trials: the CONSORT statement. *Journal of the American Medical Association*, 276, 637-639.

Berard, A., Andreu, N., Tetrault, J.P., Niyonsenga, T. and Myhal, D. (2000). Reliability of Chalmers scale to assess quality in meta-analyses on pharmacological treatments for osteoporosis. *Annals of Epidemiology*, 10(8), 498-503.

Bhogal, S.K., Teasell, R.W., Foley, N.C. and Speechley, M.R. (2005). The PEDro scale provides a more comprehensive measure of methodological quality than the Jadad scale in stroke rehabilitation literature. *Journal of Clinical Epidemiology*, 58(7), 668-673.

Boheim, R. and Taylor, M.P. (2004). Actual and preferred working hours. *British Journal of Industrial Relations*, 42(1), 149-166.

Brenner, B.M., Meyer, T.W. and Hostetter, T.H. (1982). Dietary protein intake and the progressive nature of kidney disease: The role of hemodynamically mediated glomerular injury in the pathogenesis of progressive glomerular sclerosis in aging, renal ablation, and intrinsic renal disease. *New England Journal of Medicine*, 307, 652-659.

Burke, D.G., Chilibeck, P.D., Davidson, K.S., Candow, D.G., Farthing, J. and Smith-Palmer, T. (2001). The effect of whey protein supplementation with and without creatine monohydrate combined with resistance training on lean tissue mass and muscle strength. *International Journal of Sport Nutrition and Exercise Metabolism*, 11, 349-364.

Butterfield, G.E. and Calloway, D.H. (1984). Physical activity improves protein utilisation in young men. *British Journal of Nutrition*, 51, 171-184.

Campbell, D.T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54, 297-312.

Chalmers, T.C., Smith, H. Jr. Blackburn, B. (1981). A method for assessing the quality of a randomised control trial. *Controlled Clinical Trials*, 2, 31-49.

Clark, O., Castro, A.A., Filho, J.V. and Djubelgovic, B. (2001). Interrater agreement of Jadad's scale. *Cochrane,* 1, op031.

Dempster, P. and Aitkens, S. (1995). A new air displacement method for the determination of human body composition. *Journal of Medicine and Science in Sport and Exercise*, 27, 1692-1697.

Department of Health. (1991). Dietary Reference Values for Food Energy and Nutrients for the United Kingdom (Report on Health and Social Subjects). London: HMSO.

Department of National Health and Welfare. (1990). *Nutrition Recommendations: The Report of the Scientific Review Committee*. Ottawa.

Downs, S.H. and Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health*, 52, 377-384.

Felsing, N.E., Brasel, J.A. and Cooper, D.M. (1992). Effect of low and high intensity exercise on circulating growth hormone in men. *Journal of Clinical Endocrinology and Metabolism*, 75, 157-162.

Flynn, A. (1985). Milk Protein in the Diets of those of Intermediate Years. In T.E. Galeloot and B.J. Tinbergen (Ed) *Milk Proteins ' 84*. (154-157). Wageningen, The Netherlands: Pudoc.

Food and Nutrition Board. (1989). *Recommended daily allowances*. Washington DC: National Academy.

Garn, S.M. and Kangas, J. (1981). Protein Intake, Bone Mass and Bone Loss. In H Deluca et al. (Ed). *Osteoporosis: Recent Advances in Pathogenesis and Treatment*. (257-263). University Park Press.

Goffredsen, A., Jensen, J., Borg, J. and Christiansen, C. (1986). Measurement of fat-free mass and total body fat using dual photon absorptiometry. *Metabolism*, 35, 88-93.

Hegstud, D.M. (1976). Balance studies. *Journal of Nutrition*, 106, 307-311.

Higgins, J.P.T. and Green, S. (2008). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.0* [updated February 2008]. *The Cochrane Collaboration, 2008.* Available from www.cochrane-handbook.org. 15th July 2008.

Hoffman, J.R., Ratamess, N.A., Kang, J., Falvo, M.J. and Faignebaum, A.D. (2006). Effect of protein intake on strength, body composition and endocrine changes in strength/power athletes. *Journal of the International Society of Sports Nutrition*, 3(2), 12-18.

Jadad, A., Moore, A. and Carroll, D. (1996). Assessing the quality of randomised trials: Is blinding necessary? *Controlled Clinical Trials*, 7, 1-12.

Juni, P., Altman, D.G. and Egger, M. (2001). Assessing the quality of controlled clinical trials. *British Medical Journal*, 323, 42-46.

Lachin, J.M. (1988). Properties of simple randomization in clinical trials. *Controlled Clinical Trials,* 9, 312–326.

Laritcheva, K.A., Yalovaya, N.I., Shubin, V.I. and Smirnov, P.V. (1978). Study of Energy Expenditure and Protein Needs of Top Weight-lifters. In Parazkova and Rogozkin (Ed). *Nutrition, Physical Fitness and Health*. (144-163). Baltimore: University Press.

Lemon, P.W.R. (1991). Effect of exercise on protein requirements. *Journal of Sport Science*, 9, 53-70.

Lemon, P.W.R. (1998). Effect of exercise on dietary protein requirements. *International Journal of Sports Nutrition*, 8, 426-447.

Lemon, P.W.R. (2000). Beyond the zone: Protein needs of active individuals. *Journal of the American College of Nutrition*, 19(5), 513S-521S.

Lemon, P.W.R., Tarnopolsky, M.A. and MacDougall, J.D and Atkinson, S.A. (1992). Protein requirements and muscle mass/strength changes during intensive training in novice bodybuilders. *Journal of Applied Physiology*, 73, 767-775.

Maher, C., Sherrington, C., Herbert, R., Moseley, A. and Elkins, M. (2003). Reliability of the PEDro rating quality of randomised controlled trials. *Physical Therapy*, 83, 713-721.

Manninen, A.H. (2004). High protein weight loss diets and purported adverse effects: where is the evidence? *Journal of the International Society of Sports Nutrition*, 1, 45-51.

Millward, D.J. (2001). Protein and amino acid requirements of adults: current controversies. *Canadian Journal of Applied Physiology*, 26(suppl.), S130-S140.

Oddoye, E. A. and Margen, S. (1979). Nitrogen balance studies in humans: long-term effect of high nitrogen intake on nitrogen accretion. *Journal of Nutrition*, 109, 363-377.

Olivo, S.A., Macedo, L.G., Gadotti, I.C., Fuentes, J., Stanton, T. and Magee, D.J. (2008). Scales to assess the quality of randomised controlled trials: A systematic review. *Physical Therapy*, 88(2), 156-173.

Paul, G.L. (1989). Dietary protein requirements of physically active individuals. *Journal of Sports Medicine*, 8(3), 154-176.

Phillips, S.M, Parise, G. Roy, B.D. Tipton, K.D and Wolfe, R.R. and Tamopolsky, M.A. (2002). Resistance-training-induced adaptations in skeletal muscle protein turnover in the fed state. *Canadian Journal of Physiology and Pharmacology,* 80, 1045–1053.

Phillips, S.M, Tipton, K.D, Ferrando, A.A. and Wolfe, R.R. (1999). Resistance training reduces the acute exercise-induced increase in muscle protein turnover. *American Journal of Physiology,* 276**,** E118–E124.

Rand, W.M., Pellett, P.L. and Young, V.R. (2003). Meta-analysis of nitrogen balance studies for estimating protein requirements in healthy adults. *American Journal of Clinical Nutrition*, 77(1), 109-127.

Rasch, P.J. and Pierson, W.R. (1962). Effect of protein dietary supplement on muscular strength and hypertrophy. *American Journal of Clinical Nutrition*, 11, 530-532.

Rodriguez, W.R. and Garlick, P.J. (2008). Introduction to protein summit 2007: Exploring the impact of high quality protein on optimal health. *American Journal of Clinical Nutrition*, 87(5), S1551-S1553.

Schulz, K.F. (1995). Subverting randomisation in controlled trials. *Journal of the American Medical Association*, 274, 1456-1458.

Schulz, K.F. and Grimes, D.A. (2002). Allocation concealment in randomised trials: defending against deciphering. *Lancet,* 359(9306), 614-618.

Schulz, K.F., Chalmers, I., Hayes, R.J. and Altman, D.G. (1995). Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Journal of the American Medical Association*, 273, 408-412.

Senn, S.J. (1988). Covariate imbalance and random allocation in clinical trials. *Statistics in Medicine*, 8(4), 467-475.

St. Jeor, S.T., Howard, B.V., Prewitt, E., Bovee, V., Bazzarre, T. and Eckel, R.H. (2001). Dietary protein and weight reduction. *Circulation*, 104, 1869-1874.

Steen, S.N. (1991). Precontest strategies of a male bodybuilder. *International Journal of Sports Nutrition*, 1, 69-78.

Tarnopolsky, M.A., Atkinson, S.A., MacDougall, J.D., Chesley, A., Phillips, S. and Schwarz, H.P. (1992). Evaluation of protein requirements for trained strength athletes. *Journal of Applied Physiology*, 73, 1986-1995.

Tarnopolsky, M.A., MacDougall, J.D. and Atkinson, S.A. (1988). Influence of protein intake and training status on nitrogen balance and lean body mass. *Journal of Applied Physiology*, 64(1), 187-193.

Tipton, K.D. and Wolfe, R.R. (2001). Exercise, protein metabolism, and muscle growth. *International Journal of Sports Nutrition and Exercise Metabolism*, 11(1), 109-132.

Tipton, K.D. and Wolfe, R.R. (2004). Protein and amino acids for athletes. *Journal of Sports Sciences*, 22, 65-79.

Todd, K.S., Butterfield, G.E. and Calloway, D.H. (1984). Nitrogen balance in men with adequate and deficient energy intake at three levels of work. *Journal of Nutrition*, 114(11), 2107-2118.

Tome, D. And Bos, C. (2000). Dietary protein and nitrogen utilisation. *Journal of Nutrition*, 130, 1868S-1873S.

Wilczynski, N.L. and Haynes, B.R. (2005). EMBASE search strategies for identifying methodologically sound diagnostic studies for use by clinicians and researchers. *BMC Medicine*, 3, 7-12.

Wilmore, J.H. and Costill, D.L. (2004). *Physiology of sport and exercise.* (3rd ed) Human Kinetics.

Wiseman, M.J., Hunt, R., Goodwin, A., Gross, J.L., Keen, H. And Viberti, G. (1987). Dietary composition and renal function in healthy subjects. *Nephron*, 46, 37-42.

APPENDICES

Appendix 1

List of search terms

Keywords initially used

Protein

Muscle

Strength

Specific keywords/phrases

Protein requirements

Dietary protein intake

Protein quantity

Protein supplementation

Skeletal muscle

Muscle growth

Muscle hypertrophy

Muscle mass

Strength athletes

Strength sports

Strength trained

Weightlifters

Bodybuilders

Combination of keywords/phrases

(Protein & Muscle or Strength terms)

Protein AND muscle

Protein AND skeletal muscle

Protein AND muscle growth

Protein AND muscle hypertrophy

Protein AND muscle mass

Protein AND strength

Protein AND strength athletes

Protein AND strength sports

Protein AND strength trained

Protein AND weightlifters

Protein AND bodybuilders

Appendix 2

Search strategy

(i)

The electronic databases referred to and the search terms inputted into their associated search fields

Electronic database	Specific keywords/phrases	Search fields	Combination of keywords/phrases (protein & muscle or strength)	Search fields
Blackwell Synergy	Protein requirements Dietary protein intake Protein quantity Protein supplementation Skeletal muscle Muscle growth Muscle hypertrophy Muscle mass Strength athletes Strength sports Strength trained Weightlifters Bodybuilders	Subject	Protein AND muscle Protein AND skeletal muscle Protein AND muscle growth Protein AND muscle hypertrophy Protein AND muscle mass Protein AND strength Protein AND strength athletes Protein AND strength sports Protein AND strength trained Protein AND weightlifters Protein AND bodybuilders	All fields
CINAHL Plus		Keywords		All fields
IngentaConnect		Title, keywords, abstract		Abstract/title /keywords
Pubmed		Topic		All fields
Science Direct		Title, abstract, keywords		Abstract, title, keywords
Sport Discus		Keywords		All fields

The number of search outputs generated for each search term of all six electronic databases

Blackwell Synergy

Specific keywords/phrases	Number of search outputs	Combination of keywords/phrases (protein & muscle or strength)	Number of search outputs
Protein requirements	12776	Protein AND muscle	22967
Dietary protein intake	2088	Protein AND skeletal muscle	7613
Protein quantity	1720	Protein AND muscle growth	9249
Protein supplementation	7301	Protein AND muscle hypertrophy	902
Skeletal muscle	16700	Protein AND muscle mass	2244
Muscle growth	15326	Protein AND strength	5507
Muscle hypertrophy	2021	Protein AND strength athletes	46
Muscle mass	5647	Protein AND strength sports	209
Strength athletes	246	Protein AND strength trained	35
Strength sports	1384	Protein AND weightlifters	5
Strength trained	1862	Protein AND bodybuilders	7
Weightlifters	25		
Bodybuilders	42		

CINAHL Plus

Specific keywords/phrases	Number of search outputs	Combination of keywords/phrases (protein & muscle or strength)	Number of search outputs
Protein requirements	50	Protein AND muscle	1367
Dietary protein intake	67	Protein AND skeletal muscle	349
Protein quantity	2	Protein AND muscle growth	15
Protein supplementation	51	Protein AND muscle hypertrophy	45
Skeletal muscle	1568	Protein AND muscle mass	120
Muscle growth	45	Protein AND strength	303
Muscle hypertrophy	200	Protein AND strength athletes	4
Muscle mass	640	Protein AND strength sports	1
Strength athletes	20	Protein AND strength trained	7
Strength sports	2	Protein AND weightlifters	3
Strength trained	54	Protein AND bodybuilders	8
Weightlifters	65		
Bodybuilders	62		

IngentaConnect

Specific keywords/phrases	Number of search outputs	Combination of keywords/phrases (protein & muscle or strength)	Number of search outputs
Protein requirements	1908	Protein AND muscle	9928
Dietary protein intake	1320	Protein AND skeletal muscle	2837
Protein quantity	1029	Protein AND muscle growth	77
Protein supplementation	1670	Protein AND muscle hypertrophy	48
Skeletal muscle	9414	Protein AND muscle mass	168
Muscle growth	4792	Protein AND strength	2432
Muscle hypertrophy	889	Protein AND strength athletes	3
Muscle mass	2766	Protein AND strength sports	0
Strength athletes	163	Protein AND strength trained	1
Strength sports	170	Protein AND weightlifters	0
Strength trained	285	Protein AND bodybuilders	3
Weightlifters	16		
Bodybuilders	43		

Pubmed

Specific keywords/phrases	Number of search outputs	Combination of keywords/phrases (protein & muscle or strength)	Number of search outputs
Protein requirements	28064	Protein AND muscle	87937
Dietary protein intake	17094	Protein AND skeletal muscle	28306
Protein quantity	12624	Protein AND muscle growth	593
Protein supplementation	17352	Protein AND muscle hypertrophy	353
Skeletal muscle	182082	Protein AND muscle mass	1323
Muscle growth	70061	Protein AND strength	14725
Muscle hypertrophy	17723	Protein AND strength athletes	10
Muscle mass	36321	Protein AND strength sports	1
Strength athletes	1268	Protein AND strength trained	20
Strength sports	5900	Protein AND weightlifters	10
Strength trained	1699	Protein AND bodybuilders	24
Weightlifters	140		
Bodybuilders	215		

Science Direct

Specific keywords/phrases	Number of search outputs	Combination of keywords/phrases (protein & muscle or strength)	Number of search outputs
Protein requirements	7586	Protein AND muscle	10441
Dietary protein intake	2609	Protein AND skeletal muscle	2743
Protein quantity	5638	Protein AND muscle growth	2038
Protein supplementation	2795	Protein AND muscle hypertrophy	270
Skeletal muscle	18042	Protein AND muscle mass	782
Muscle growth	7897	Protein AND strength	2808
Muscle hypertrophy	1623	Protein AND strength athletes	12
Muscle mass	4704	Protein AND strength sports	12
Strength athletes	186	Protein AND strength trained	6
Strength sports	256	Protein AND weightlifters	0
Strength trained	504	Protein AND bodybuilders	6
Weightlifters	20		
Bodybuilders	66		

Sport Discus

Specific keywords/phrases	Number of search outputs	Combination of keywords/phrases (protein & muscle or strength)	Number of search outputs
Protein requirements	94	Protein AND muscle	2999
Dietary protein intake	22	Protein AND skeletal muscle	836
Protein quantity	1	Protein AND muscle growth	103
Protein supplementation	58	Protein AND muscle hypertrophy	76
Skeletal muscle	4708	Protein AND muscle mass	239
Muscle growth	443	Protein AND strength	628
Muscle hypertrophy	278	Protein AND strength athletes	12
Muscle mass	1428	Protein AND strength sports	0
Strength athletes	103	Protein AND strength trained	11
Strength sports	32	Protein AND weightlifters	16
Strength trained	139	Protein AND bodybuilders	368
Weightlifters	766		
Bodybuilders	5823		

(ii)

Grey literature

Number of search outputs generated from each search term on the online search engine, 'Scirus'

Specific keywords/phrases	Number of search outputs	Combination of keywords/phrases (protein & muscle or strength)	Number of search outputs
Protein requirements	17	Protein AND muscle	168
Dietary protein intake	2	Protein AND skeletal muscle	43
Protein quantity	9	Protein AND muscle growth	6
Protein supplementation	23	Protein AND muscle hypertrophy	0
Skeletal muscle	638	Protein AND muscle mass	6
Muscle growth	107	Protein AND strength	23
Muscle hypertrophy	10	Protein AND strength athletes	0
Muscle mass	100	Protein AND strength sports	0
Strength athletes	10	Protein AND strength trained	0
Strength sports	4	Protein AND weightlifters	0
Strength trained	2	Protein AND bodybuilders	0
Weightlifters	6		
Bodybuilders	7		

Hand searching, Peak Performance newsletter

Number of search outputs (from 17 references)
2

The initial search terms within the title of the relevant references deriving from this source

	Initial search terms in title
Reference 1	Protein, Muscle, Strength
Reference 2	Protein, Muscle

(iii)

Reference Chasing

Number of search outputs (from 4 studies included)
8

Number of references deriving from the reference lists of each of the four studies

		Initial search terms in title
Hoffman, Ratamess, Kang, Falvo & Faigenbaum, (2006)	Reference 1	Protein, Muscle
	Reference 2	Protein, Strength
	Reference 3	Protein, Muscle
Lemon , Tarnopolsky, MacDougall & Atkinson, (1992)	Reference 4	Protein, Strength
	Reference 5	Protein, Muscle
	Reference 6	Protein, Muscle
Tarnopolsky, MacDougall & Atkinson, (1988)	Reference 7	Protein, Strength
Tarnopolsky *et al*, (1992)	Reference 8	Protein, Strength

Appendix 3

Origin of included studies and search term (s) that generated reference after the completion of the assessment of 'immediate' relevance

Source of reference	Search term reference derives from
CINAHL Plus	Reference 1 & 2- 'Muscle hypertrophy' Reference 3- 'Bodybuilders' Reference 4- 'Protein & Skeletal muscle' Reference 5- 'Protein & Strength trained'
Pubmed	Reference 6- 'Protein & Strength athletes' Reference 7 & 8- 'Protein & Bodybuilders'
Sports Discus	Reference 9- 'Protein supplementation' Reference 10- 'Muscle mass'
'Peak Performance' newsletter	Reference 11- 'Protein', 'Muscle', 'Strength' Reference 12- 'Protein', 'Muscle'

Appendix 4

Articles excluded on basis of *Orientation of study*-

Effect of total protein intake on muscle (n=6)

Full reference	Source	Reason for exclusion
Bamman, M.M., Hunter, G.R., Newton, L.E., Roney, R.K & Khaled, M.A. (1993). Changes in body composition, diet, and strength of bodybuilders during the 12 weeks prior to competition. *Journal of Sports Medicine and Physical Fitness*, 33(4), 383-391.	Electronic database	Excluded on comprehensive relevance criteria-did not show effect of total protein intake on muscle
Colker, C.M., Swain, M.A., Fabrucini, B., Shi, Q. & Kalman, D.S. (2000). Effects of supplemental protein on body composition and muscular strength in healthy athletic male adults. *Current Therapeutic Research*, 61(1), 19-28.	Reference chasing	Excluded on comprehensive relevance criteria-did not show effect of total protein intake on muscle
Cribb, P.J. & Hayes, A. (2006). Effects of supplement timing and resistance exercise on skeletal muscle hypertrophy. *Journal of Medicine and Science in Sport and Exercise*, 38(11), 1918-1925.	Electronic database	Excluded on comprehensive relevance criteria-did not show effect of total protein intake on muscle
Tarnopolsky, M.A., Lemon, P.W.R., MacDougall,	Reference	Excluded on comprehensive

J.D. & Atkinson, S.A. (1990). Effect of bodybuilding exercise on protein requirements (Abstract). *Canadian Journal of Applied Sport Science*, 15, 22S.	chasing	relevance criteria-did not show effect of total protein intake on muscle
Van Zant, R.S., Conway, J.M. & Seale, J.L. (2002). A moderate carbohydrate and fat diet does not impair strength performance in moderately trained males. *Journal of Sports Medicine and Physical Fitness*, 42(1), 31-37.	Electronic database	Excluded on comprehensive relevance criteria-did not show effect of total protein intake on muscle
Walberg, R.L., Leidy, M.K., Sturgill, D.J., Hinkle, D.E., Ritchey, S.J. & Sebolt, D.R. (1988). Macronutrient content of a hypoenergy diet affects nitrogen retention and muscle function in weight lifters. *International Journal of Sports Medicine*, 9(4), 261-266.	Electronic database	Excluded on comprehensive relevance criteria-did not show effect of total protein intake on muscle

Articles excluded on basis of *Methodological design-*

Comparison protein treatment group (s) (n=4)

Full reference	Source	Reason for exclusion
Cribb, P.J., Williams, A.D., Stathis, C.G., Carey, M.F. & Hayes, A. (2007). Effects of whey isolate, creatine, and resistance training on muscle hypertrophy. *Journal of Medicine and Science in Sports and Exercise*, 39(2), 298-307.	Electronic database	Excluded on comprehensive relevance criteria- no comparison protein treatment group
Cribb, P.J., Williams, A.D. & Hayes, A. (2007). A creatine-protein-carbohydrate supplement enhances responses to resistance training. *Journal of Medicine and Science in Sports and Exercise*, 39(11), 1960-1968.	Electronic database	Excluded on comprehensive relevance criteria- no comparison protein treatment group
Hartman, J.W., Tang, J.E., Wilkinson, S.B., Tarnopolsky, M.A., Lawrence, R.L. Fullerton, A.V & Phillips, S.M. (2007). Consumption of fat-free fluid milk after resistance exercise promotes greater lean mass accretion than does consumption of soy or carbohydrate in young, novice, male weightlifters. *American Journal of Clinical Nutrition*, 86(2), 373-381.	Electronic database	Excluded on comprehensive relevance criteria- no comparison protein treatment group
Tang, J.E., Manolakos, J.J., Lysecki, P.J., Moore,	Electronic	Excluded on comprehensive

D.R. & Phillips, S.M. (2007). Minimal whey protein with carbohydrate stimulates muscle protein synthesis following resistance exercise in trained young men. *Journal of Applied Physiology, Nutrition and Metabolism*, 32(6), 1132-1137.	database	relevance criteria- no comparison protein treatment group

Appendix 5

Hoffman, Ratamess, Kang, Falvo & Faignebaum, (2006).

Research design	Open controlled counterbalanced repeated measures protocol
Sample	23 male collegiate strength/power athletes aged between 18-24 years
Treatment groups	Subjects assigned to 1 of 3 separate groups: Below recommended daily protein intake (BL)-1.0-1.4g/kg Bw/day Recommended daily protein intake (RL)-1.6-1.8g/kg Bw/day Above recommended daily protein intake (AL)->2.0g/kg Bw/day
Interventions	12 weeks of same resistance training program, (4 days per week split routine) and at the same time maintaining the assigned daily protein intake. Testing of lean body mass (kg) occurred prior to onset of training program (PRE) and after its conclusion (POST). Calorie intake was similar across all treatment groups (3100-3200 kcals).
Statistical analysis	A critical alpha level of $p \leq 0.05$ was used to determine statistical significance
Outcome measure	1.4g/kg Bw/day-no muscle growth occurred above this quantity (i.e. 1.0-1.4g/kg Bw/day or BL amount)

Lemon, Tarnopolsky, MacDougall & Atkinson, (1992)

Research design	Randomised double blind counterbalanced crossover protocol
Sample	14 male novice bodybuilders aged between 20-25 years
Treatment groups	Subjects randomly assigned to 2 separate groups: Total protein intake of 2.62g/kg Bw/day Total protein intake of 1.35g/kg Bw/day
Interventions	Two 1 month dietary treatment groups separated by a 7 day ad libitum diet washout period. During the 8 weeks undertaking both treatment groups, an intensive weight training program was undertaken 6 days/week. Energy compositions of both treatment groups are similar, (4000-4100 kcals).
Statistical analysis	$p \leq 0.05$ was taken to indicate significance.
Outcome measure	1.35g/kg Bw/day as no further muscle growth occurs above this quantity.

Tarnopolsky, MacDougall & Atkinson, (1988)

Research design	Open controlled crossover protocol
Sample	6 male elite bodybuilders aged 23-25 years
Treatment groups	Subjects randomly assigned to 2 separate groups: Exp A. Normal diet-Total protein intake of 2.77g/kg Bw/day Exp B. Altered diet-Total protein intake of 1.05g/kg Bw/day
Interventions	13 days were assigned to each experiment. During the 13 days their normal 3 day split routine of resistance training was performed. Energy intake was maintained across both treatment groups, (4800 kcals).
Statistical analysis	A confidence level of $p \leq 0.05$ was taken to indicate significance
Outcome measure	1.05g/kg Bw/day as no significant increase in lean body mass with 2.77g/kg Bw/day

Tarnopolsky *et al*, (1992)

Research design	Randomised counterbalanced crossover protocol
Sample	2 groups: 7 male strength athletes aged between 20-23 years 6 healthy sedentary males in control group aged between 20-28 years
Treatment groups	Both groups randomly assigned to 3 separate groups: Low dietary protein intake (LP)-0.86g/kg Bw/day Moderate protein intake (MP)-1.4g/kg Bw/day High protein intake (HP)-2.4g/kg Bw/day
Interventions	13 day experimental periods with an 8 day washout period involving diet allocation and form of resistance training, primarily circuit weight routines. Calorie intake was similar across all treatment groups. Sedentary group (2500-2600 kcals) have reduced calorie intake compared to strength athletes (3600-3700 kcals).
Statistical analysis	$p \leq 0.05$ was taken to indicate significance
Outcome measure	1.4g/kg Bw/day as rate of amino acid oxidation increased over this quantity and muscle mass did not significantly increase, signifying that the maximum protein intake lies here

Appendix 6

Checklist derives from Downs and Black, (1998)

Appendix

Checklist for measuring study quality

Reporting

1. *Is the hypothesis/aim/objective of the study clearly described?*

yes	1
no	0

2. *Are the main outcomes to be measured clearly described in the Introduction or Methods section?*
If the main outcomes are first mentioned in the Results section, the question should be answered no.

yes	1
no	0

3. *Are the characteristics of the patients included in the study clearly described ?*
In cohort studies and trials, inclusion and/or exclusion criteria should be given. In case-control studies, a case-definition and the source for controls should be given.

yes	1
no	0

4. *Are the interventions of interest clearly described?*
Treatments and placebo (where relevant) that are to be compared should be clearly described.

yes	1
no	0

5. *Are the distributions of principal confounders in each group of subjects to be compared clearly described?*
A list of principal confounders is provided.

yes	2
partially	1
no	0

6. *Are the main findings of the study clearly described?*
Simple outcome data (including denominators and numerators) should be reported for all major findings so that the reader can check the major analyses and conclusions. (This question does not cover statistical tests which are considered below).

yes	1
no	0

7. *Does the study provide estimates of the random variability in the data for the main outcomes?*
In non normally distributed data the inter-quartile range of results should be reported. In normally distributed data the standard error, standard deviation or confidence intervals should be reported. If the distribution of the data is not described, it must be assumed that the estimates used were appropriate and the question should be answered yes.

yes	1
no	0

8. *Have all important adverse events that may be a consequence of the intervention been reported?*
This should be answered yes if the study demonstrates that there was a comprehensive attempt to measure adverse events. (A list of possible adverse events is provided).

yes	1
no	0

9. *Have the characteristics of patients lost to follow-up been described?*
This should be answered yes where there were no losses to follow-up or where losses to follow-up were so small that findings would be unaffected by their inclusion. This should be answered no where a study does not report the number of patients lost to follow-up.

yes	1
no	0

10. *Have actual probability values been reported(e.g. 0.035 rather than <0.05) for the main outcomes except where the probability value is less than 0.001?*

yes	1
no	0

External validity

All the following criteria attempt to address the representativeness of the findings of the study and whether they may be generalised to the population from which the study subjects were derived.

11. *Were the subjects asked to participate in the study representative of the entire population from which they were recruited?*
The study must identify the source population for patients and describe how the patients were selected. Patients would be representative if they comprised the entire source population, an unselected sample of consecutive patients, or a random sample. Random sampling is only feasible where a list of all members of the relevant

population exists. Where a study does not report the proportion of the source population from which the patients are derived, the question should be answered as unable to determine.

yes	1
no	0
unable to determine	0

12. *Were those subjects who were prepared to participate representative of the entire population from which they were recruited?*
The proportion of those asked who agreed should be stated. Validation that the sample was representative would include demonstrating that the distribution of the main confounding factors was the same in the study sample and the source population.

yes	1
no	0
unable to determine	0

13. *Were the staff, places, and facilities where the patients were treated, representative of the treatment the majority of patients receive?*
For the question to be answered yes the study should demonstrate that the intervention was representative of that in use in the source population. The question should be answered no if, for example, the intervention was undertaken in a specialist centre unrepresentative of the hospitals most of the source population would attend.

yes	1
no	0
unable to determine	0

Internal validity - bias

14. *Was an attempt made to blind study subjects to the intervention they have received ?*
For studies where the patients would have no way of knowing which intervention they received, this should be answered yes.

yes	1
no	0
unable to determine	0

15. *Was an attempt made to blind those measuring the main outcomes of the intervention?*

yes	1
no	0
unable to determine	0

16. *If any of the results of the study were based on "data dredging", was this made clear?*
Any analyses that had not been planned at the outset of the study should be clearly indicated. If no retrospective unplanned subgroup analyses were reported, then answer yes.

yes	1
no	0
unable to determine	0

17. *In trials and cohort studies, do the analyses adjust for different lengths of follow-up of patients, or in case-control studies, is the time period between the intervention and outcome the same for cases and controls ?*
Where follow-up was the same for all study patients the answer should yes. If different lengths of follow-up were adjusted for by, for example, survival analysis the answer should be yes. Studies where differences in follow-up are ignored should be answered no.

yes	1
no	0
unable to determine	0

18. *Were the statistical tests used to assess the main outcomes appropriate?*
The statistical techniques used must be appropriate to the data. For example non-parametric methods should be used for small sample sizes. Where little statistical analysis has been undertaken but where there is no evidence of bias, the question should be answered yes. If the distribution of the data (normal or not) is not described it must be assumed that the estimates used were appropriate and the question should be answered yes.

yes	1
no	0
unable to determine	0

19. *Was compliance with the intervention/s reliable?*
Where there was non compliance with the allocated treatment or where there was contamination of one group, the question should be answered no. For studies where the effect of any misclassification was likely to bias any association to the null, the question should be answered yes.

yes	1
no	0
unable to determine	0

20. *Were the main outcome measures used accurate (valid and reliable)?*

For studies where the outcome measures are clearly described, the question should be answered yes. For studies which refer to other work or that demonstrates the outcome measures are accurate, the question should be answered as yes.

yes	1
no	0
unable to determine	0

Internal validity - confounding (selection bias)

21. *Were the patients in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited from the same population?*
For example, patients for all comparison groups should be selected from the same hospital. The question should be answered unable to determine for cohort and case-control studies where there is no information concerning the source of patients included in the study.

yes	1
no	0
unable to determine	0

22. *Were study subjects in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited over the same period of time?*
For a study which does not specify the time period over which patients were recruited, the question should be answered as unable to determine.

yes	1
no	0
unable to determine	0

23. *Were study subjects randomised to intervention groups?*
Studies which state that subjects wererandomised should be answered yes except where method of randomisation would not ensure random allocation. For example alternate allocation would score no because it is predictable.

yes	1
no	0
unable to determine	0

24. *Was the randomised intervention assignment concealed from both patients and health care staff until recruitment was complete and irrevocable?*
All non-randomised studies should be answered no. If assignment was concealed from patients but not from staff, it should be answered no.

yes	1
no	0
unable to determine	0

25. *Was there adequate adjustment for confounding in the analyses from which the main findings were drawn?*
This question should be answered no for trials if: the main conclusions of the study were based on analyses of treatment rather than intention to treat; the distribution of known confounders in the different treatment groups was not described; or the distribution of known confounders differed between the treatment groups but was not taken into account in the analyses. In non-randomised studies if the effect of the main confounders was not investigated or confounding was demonstrated but no adjustment was made in the final analyses the question should be answered as no.

yes	1
no	0
unable to determine	0

26. *Were losses of patients to follow-up taken into account?*
If the numbers of patients lost to follow-up are not reported, the question should be answered as unable to determine. If the proportion lost to follow-up was too small to affect the main findings, the question should be answered yes.

yes	1
no	0
unable to determine	0

Power

27. *Did the study have sufficient power to detect a clinically important effect where the probability value for a difference being due to chance is less than 5%?*
Sample sizes have been calculated to detect a difference of x% and y%.

	Size of *smallest* intervention group	
A	$<n_1$	0
B	n_1–n_2	1
C	n_3–n_4	2
D	n_5–n_6	3
E	n_7–n_8	4
F	n_9+	5

APPENDIX 7

Scoring of studies using Downs and Black checklist for quality assessment

		Downs & Black checklist																													
		REPORTING (10 items)										EXTERNAL VALIDITY (3 items)			INTERNAL VALIDITY (bias and selection bias) (13 items)													POWER (1 item)		
	MAXIMUM SCORE FOR EACH ITEM	**1**	**1**	**1**	**1**	**2**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**5**	**TOTAL SCORE /31**	**PASS SCORE FOR INCLUSION IN REVIEW, 75% OR 23/31**
STUDY	Hoffman *et al*, (2006)	1	1	1	1	0	1	1	1	1	0	1	1	1	0	0	1	1	1	1	1	1	1	1	0	0	0	1	5	24	PASS
	Lemon *et al*, (1992)	1	1	1	1	0	1	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	3	24	PASS
	Tarnopolsky *et al*, (1988)	1	1	1	1	0	1	1	0	1	1	0	0	1	0	0	1	1	1	1	1	1	1	1	0	0	0	1	3	20	FAIL
	Tarnopolsky *et al*, (1992)	1	1	1	1	0	1	1	1	1	1	1	0	1	0	0	1	1	1	1	1	1	1	1	1	1	0	1	4	25	PASS

(Refer to Appendix 6 for questions associated with each item)

KEY:

RED text= non-randomised study

BLACK text= randomised study