Department of Clinical Sciences & Nutrition

MSc

Human Nutrition

Project Title	The effect of glycomacropeptide-based foods upon blood phenylalanine control in adults and children with phenylketonuria.

Module title	Research Project
Module code	XN7006
Supervisor	Dr Sohail Mushtaq

Student J number	J31913
Year of intake	2016
Date submitted	03/ 09/ 2018

Word count	
Review Paper	**5051**
Research Article	**6384**

Table of contents

Section ***Pages***

Table of contents

Section ***Pages***

Table of contents (continued)

Section ***Pages***

List of tables and figures in literature review

Table/ Figure ***Page***

List of tables and figures in systematic review

Student declaration.

I (Mr Roderick Thomson) declare that this assignment is my own work and that I have acknowledged the work of others in accordance with University and School guidance on good academic conduct and how to avoid plagiarism.

Glossary of abbreviations and terms used.

Phe (phenylalanine): An essential amino acid that patients with PKU are unable to metabolise normally.

PKU (Phenylketonuria): An autosomal recessive, inborn error of metabolism affecting the hepatic enzyme PAH that causes phenylalanine (from diet or catabolism) to accumulate to toxic levels, causing neurocognitive defects.

PAH (Phenylalanine Hydroxylase): Hepatic enzyme, the functioning of which is impaired in PKU due to mutations in the pair of alleles coding for it. This causes Phe to accumulate and makes Tyrosine essential as PAH ordinarily converts phe to it.

BH_4 (biopterin): co-factor required for PAH function. Mutations affecting enzymes involved in biosynthesis and regeneration impair PAH and cause Phe accumulation resembling PKU, but BH_4 defects require differential treatment. Synthetic BH_4 (Sapropterin) is a novel, oral treatment for PKU which improves PAH function in some cases, but only subset of less severe cases (with some residual PAH functioning).

Plasma phe: used to describe phe levels in blood, which are elevated in PKU and closely predict the extent of neurocognitive impairments. PKU is detected through elevations in plasma phe and treatment aims to control phe to avoid symptoms.

Control: the objective of treatment is to 'control' phe. Treatment guidelines recommend maintaining phe within ranges associated with neurocognitive outcomes most similar to healthy controls.

Glossary of abbreviations and terms used (continued).

Neurocognitive defects: deficits in (for example) intelligence, executive functions (e.g. reaction times) and movement disorders which are severe enough to necessitate lifelong institutionalisation in untreated PKU. However outcomes are often below-normal even in early-treated patients. This partly explains interest in novel treatments.

Dietary treatment: Is initiated at birth to control plasma phe by restricting phe intake, which involves restricting most protein-containing foods, even staples. Phe-free protein substitutes (formula) are thus also needed to meet protein requirements.

Early treated: Patients treated within 3 months of birth. This is crucial as the developing CNS is particularly vulnerable to elevated phe and defects are worse if PKU is 'late-treated' (from 3 months to 7 years) and severely disabling if 'untreated' (until after 7 years). For the same reason, control targets are stricter (and tolerance lower) in children (until ~12-18 years of age depending on the guideline used).

Formula: phe-free protein substitutes. These are unpopular due to their poor taste, social acceptability and inconvenience which likely explain many patients struggling to comply with treatment. They are improving however, as flavoured, ready-to-drink pouches are now available whilst powdered versions requiring weighing and mixing were once the norm.

Compliance with therapeutic diets: involves consuming sufficient formula, distributing it over 3-4 meals and eating the correct amounts of dietary protein and energy. Non-compliance alters control by increasing phe intake or affecting the balance of anabolism (which lowers phe by incorporating it into tissues) and catabolism (which does the opposite).

Tolerance: Amount of dietary phe a patient can consume without losing control. Influenced by severity I.E. the extent to which PAH functions, which is determined by the mutation pair inherited. However not all are characterised and tolerance is not always predictable from genetic tests. Tolerance is worked out in practice by monitoring control and adjusting diet.

Exchanges: System used to communicate tolerance to patients in food terms. Patients are assigned a number of exchanges of protein per day depending on tolerance and documentation lists how many exchanges foods contain. One exchange assumes 50mg phe per g protein which is a slight simplification.

Guthrie Assay: The landmark, semi-quantitative method used to detect PKU at birth, allowing immediate treatment. Now superseded by more reliable methods such as tandem mass spectrometry of dried blood spots (DBS).

Glycomacropeptide (GMP): The only naturally Phe-free intact protein. Used to manufacture medical foods forming novel alternatives to formula. These have several theoretical advantages over formula, particularly their acceptability. However, commercially viable purification processes are imperfect leaving GMP contaminated with some residual phe, raising concerns around its effect on control.

LNAA (Large Neutral Amino Acids): A group of amino acids including phe which are absorbed from the gastrointestinal tract and cross the blood brain barrier via a shared transporter (L-type). LNAA treatment appears to reduce Plasma phe by competitively inhibiting phe uptake. GMP is naturally high in two LNAA (Thr & Ile) but requires supplementation with others (Arg, His, Leu, Trp, Tyr) to form a complete amino acid source. The degree of supplementation may thus influence GMPs impact on control

The effect of glycomacropeptide-based foods upon blood phenylalanine control in adults and children with phenylketonuria: A non-systematic literature review.

0.0: Abstract.

Conventional treatment for phenylketonuria restricts dietary phenylalanine to 'control' plasma phenylalanine concentrations. Its widespread adoption has largely eradicated the severe neurocognitive defects that previously characterised phenylketonuria. However, interest in alternative treatments continues as deficits in intelligence and other health outcomes remain problematic, conventional treatment has limitations and adherence proves difficult. Glycomacropeptide-based foods (GMP) are a novel treatment that may improve the satiety and acceptability of dietary treatment and address suboptimal health outcomes. However, glycomacropeptide contains some phenylalanine, raising safety concerns regarding its effect on plasma phenylalanine in adults and particularly children who tolerate less phenylalanine. This narrative review attempted to resolve these concerns. Its findings suggest adults and children can maintain control on GMP but individualised titrations, adjusting the amount of GMP consumed whilst monitoring plasma phenylalanine, are necessary in children. Equivalent control is a supportive finding given GMPs many advantages but this must be viewed cautiously as only seven studies were located, predominantly employing bias-prone, heterogeneous designs. GMPs effect upon control thus requires clarification via a systematic review using evidence-based, transparent methods to synthesize the entire evidence base and consider the impact of design quality, bias and heterogeneity upon results.

1.0: Introduction.

This narrative review evaluates the utility and safety of glycomacropeptide, a novel treatment for phenylketonuria. It first establishes phenylketonuria's metabolic basis and how plasma phenylalanine concentrations, which predict the extent of neurocognitive symptoms, form the central marker in its study and treatment. It then describes the

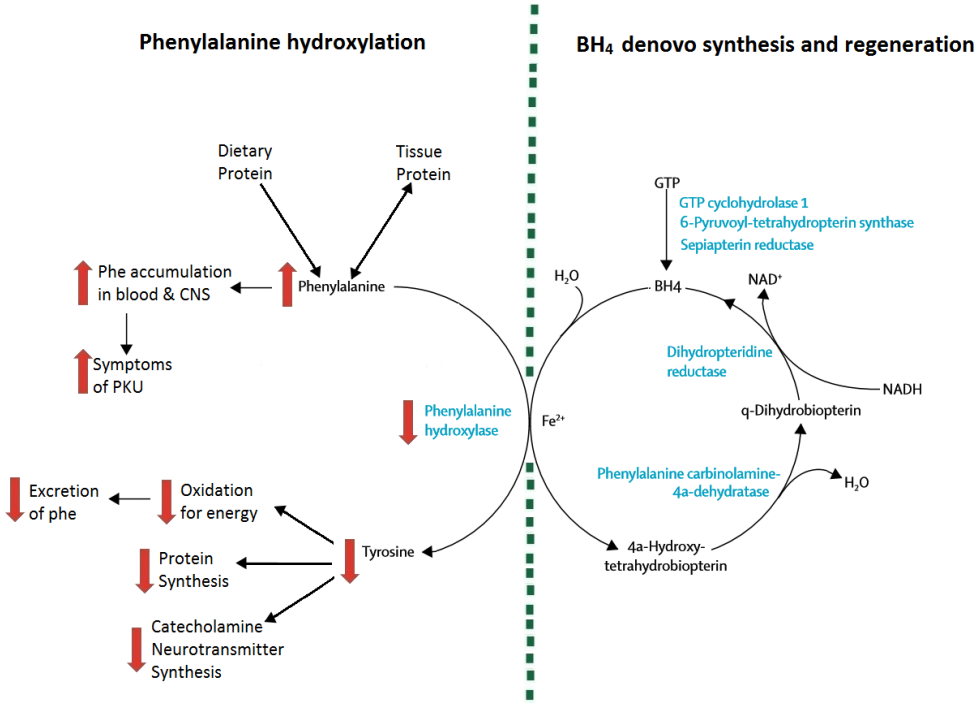
severely disabling neurocognitive impairments that characterised phenylketonuria before treatment was discovered. Dietary treatment is next outlined. This restricts phenylalanine intake to control plasma phenylalanine concentrations (henceforth: plasma phe) and its widespread adoption has largely eradicated the life-changing disabilities that were previously commonplace. Despite such improvements, it is argued that intelligence and other health outcomes often remain below-normal, that treatment is burdensome and difficult to adhere to and that both factors explain continuing interest in novel treatments. Glycomacropeptide, one such treatment, is next evaluated. It is concluded that glycomacropeptide has the potential to improve the acceptability of therapeutic diets and ameliorate other suboptimal health outcomes, despite a lack of longer-term outcome data. However, glycomacropeptide contains phenylalanine, raising safety concerns regarding its impact on plasma phenylalanine control and neurocognitive outcomes. The reviews findings suggest maintenance of plasma phenylalanine is possible in adults and in children, provided individualised adjustments are performed in children but this conclusion is tentative as included studies predominantly employed bias-prone, heterogeneous designs. Resultantly, the review recommends that GMPs effect upon control is clarified using systematic methods that synthesize the entire evidence base and consider the impact of bias and heterogeneity upon results.

2.0: Metabolic basis of Phenylketonuria.

Phenylketonuria (PKU) is an autosomal recessive, inborn error of metabolism (IEM) that impairs functionality of the hepatic enzyme Phenylalanine Hydroxylase (PAH). PAH physiologically hydroxylates the essential amino acid (AA) phenylalanine, from dietary protein or tissue catabolism, to tyrosine (figure 1, left) (Williams, Mamotte, & Burnett, 2008). Tyrosine is used biosynthetically or oxidised (Blau, van Spronsen & Levy, 2010) and

excess phenylalanine is excreted via tyrosine oxidation so its suppression (by decreased PAH activity) in Phenylketonuria causes phenylalanine accumulation in blood (HPA) and tissues (Blau et al., 2010; Williams et al., 2011).

Figure 1. Suppression of the hydroxylation and excretion of Phe in PKU (left) and dependence of PAH on BH_4 synthesis and regeneration (right).



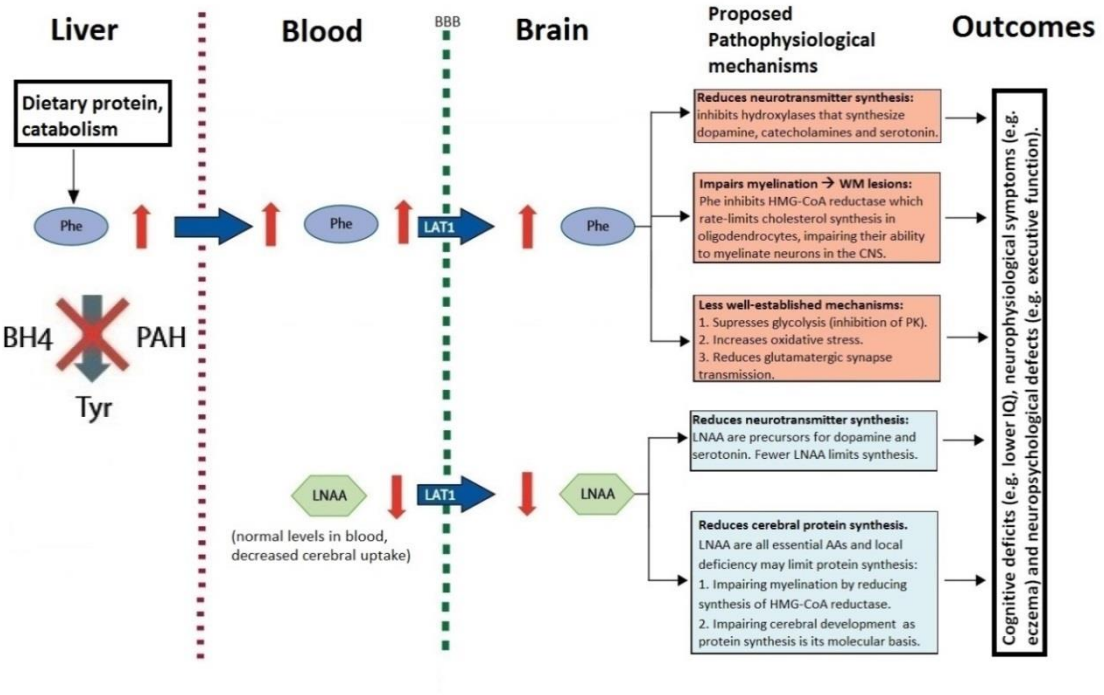
CNS = central nervous system. GTP = guanosine triphosphate. NADH = Nicotinamide adenine dinucleotide. Adapted from Blau et al., (2010).

Tetrahydrobiopterin (BH_4) is a co-factor required by PAH. Mutations affecting enzymes involved in BH_4 synthesis and regeneration (figure 1, right) can limit PAH activity, causing HPA resembling PKU (Longo, 2009). However BH_4 has other roles (e.g. neurotransmitter synthesis) so BH_4 defects require differential diagnosis and treatment (Blau, Hennermann, Langenbeck & Lichter-Konecki, 2011) and are henceforth excluded from the term PKU.

2.1: Pathophysiology of PKU and importance of plasma phe.

How phe accumulation causes the neurocognitive symptoms that explain most morbidity in PKU (section 1.2) remains unclear. A detailed review of hypothetical mechanisms, which are not mutually exclusive, is unfeasible (see de Groot, Hoeksma, Blau, Reijngoud & van Spronsen, 2010; Blau et al., 2010) but these are summarised in figure 2 and two points noted.

Figure 2: Mechanisms potentially linking Phe accumulation to neurocognitive symptoms.



Tyr = Tyrosine. BBB = Blood-Brain Barrier. LNAA = Large Neutral Amino Acids. LAT1 = LNAA type 1 transporter. HMG-CoA Reductase = Hydroxy-methyl-glutaryl-coenzyme-A reductase. PK = Pyruvate Kinase. Adapted from van Spronsen et al., (2017).

Firstly, two disturbances (secondary to HPA) initiate every mechanism: elevated cerebral phe concetrations and decreased cerebral LNAA concentrations. Both occur as phe and other LNAA cross the BBB via the LAT-1 counter-transporter and HPA favours Phe uptake,

which competitively inhibits LNAA uptake and promotes LNAA counter-efflux (Pietz et al., 1999; de Groot et al., 2010). Most importantly, the only factor implicated in every mechanism, routinely measurable (van Spronsen et al., 2017) and robustly predictive of neurocognitive outcomes is plasma phe as the extent and duration of elevations predict the severity of neurocognitive defects (Waisbren et al., 2007; Fonnesbeck, McPheeters, Krishnaswami, Lindegren & Reimschisel, 2012). Therefore, whatever pathological mechanism(s) plasma phe concentration forms a surrogate for, it is the central marker in the study and treatment of PKU. Elevations are used to diagnose PKU and treatment essentially maintains concentrations linked to favourable neurocognitive outcomes in the literature (van Wegberg et al., 2017).

2.2: PKU is rare but causes severe disability if untreated.

Compared to diseases considered major burdens to healthcare like diabetes PKU is rare.

Table 1. Prevalence of PKU in UK with international comparisons

Disease and Population	**Estimated Prevalence**	**Possible explanation**	**Source of estimate**
Diabetes in England (over 16s only)	***1 in 11.6***	--	*Public Health England (2016)*
PKU in Europe	***1 in 10,000***	--	*Loeber et al., (2001)*
PKU in Turkey (incidence)	1 in 4000	Consanguineous marriage	Williams et al., (2008)
PKU in Ireland	1 in 6200	Historically small & isolated gene pool	Loeber et al., (2001)
PKU in Scotland	1 in 7802	As above	Loeber et al., (2001)
PKU in Caucasians living in England	1 in 10,000	--	Hardelid et al., (2008)
PKU in Wales	1 in 10,700	--	Loeber et al., (2001)
PKU in South Asians living in England	1 in 33,000	Founder effects and/ or genetic drift	Hardelid et al., (2008)
PKU in Sub-Saharan Africans in England	1 in 100,000	As above	Hardelid et al., (2008)
Finland (incidence)	1 in 200,000	As above	Williams et al., (2008)
BH_4 Defects	<2% of PAH cases		Longo (2009)

Adapted from essay previously submitted.

Untreated PKU is nevertheless associated with severe disability. Case reviews and surveys pre-dating the proliferation of treatment (Jervis, 1937; Paine, 1957) or involving patients born before it (Murphy et al., 2008; Mazur et al., 2011) describe severe intellectual impairments in most patients: ~70% attain IQs under 20 versus 1990 norms of ~110 (Smith, Beasley & Ades, 1990) and lifelong, 24-hour care is usually necessary.

3.0: Treatment from birth and strict control in childhood are essential.

Dietary treatment restricts phe intake to maintain plasma phe within ranges linked to neurocognitive outcomes (particularly intelligence and executive functions) closest to healthy controls in studies, particularly large syntheses (Waisbren et al., 2007; Fonnesbeck et al., 2012; Albrecht, Garbade & Burgard, 2009). This provides sufficient phe for growth and protein turnover but avoids accumulation. Treatment was discovered in 1954 (Bickel, Gerrard & Hickmans, 1954; Bickel, 1996). Those initially treated enjoyed marked intelligence improvements over untreated norms (MRC, 1963) but the inability to reliably diagnose PKU in newborns delayed treatment (Guthrie & Susi, 1963) despite the belief that immediate treatment would protect the rapidly developing (and theoretically more sensitive) CNS from phe and further improve outcomes (MRC, 1963; Guthrie & Susi, 1963; Koch & de la Cruz, 1999). The Guthrie test was later developed (Guthrie & Susi, 1963) permitting screening and treatment at birth which rolled out across the USA and UK from ~1964 - 1971 (Koch & de la Cruz, 1999; Smith et al., 1990). Nationwide prospective studies during the rollout noted stratified associations between the rapidity of treatment initiation, IQ and school performance in ages 4-12, confirming the importance of early treatment (Smith et al., 1990; Beasley et al., 1994; Azen et al., 1991; Azen, Koch, Friedman, Wenz & Fishler, 1996). On this (observational) basis, newborn screening and early treatment are universally recommended and practiced in most developed nations, though policies vary (table 2).

Table 2 : Screening policies, diagnostic thresholds and recommended control ranges

Treatment Guideline or policy (and source)	**Recommendations for screening and treatment initiation at birth.**	**Threshold used to diagnose PKU at birth (untreated plasma phe in µmol/ L)**	**Recommended plasma phe ranges in early life µmol/L**	**Recommended plasma phe ranges in later life (µmol/L)**	**Notes**
ESPKU European Guidelines (van Wegberg, 2017)	Screen by day 3. Treat by day 10.	>360 µmol/ L	120 – 360 Until age 12	120 – 600 12 onwards (lifelong)	Extent to which guidelines will be implemented in each European state is unclear.
National Institutes of Health USA guidelines (NIH CDP, 2001)	Universal screening since ~1963-70. Screen ASAP. Treat by day 7-10.	>360 µmol/ L	120 – 360 Until age 12	120 – 900 12 onwards (lifelong)	Implementation varies by state/ clinic.
MRC UK Guidelines (Smith et al, 1993)	Universal screening since ~1971. Screen by day 5. Treat by day 20.	>400 µmol/ L	120 – 360 Until age 6 120 – 480 Ages 6 - 16	120 – 700 16 onwards (lifelong)	Likely to be replaced by more recent ESPKU guidelines.
German Guidelines (Burgard et al., 1999)	Universal screening is standard. Screen and treat as soon as possible.	>600 µmol/ L	40 – 240 Until age 10 40 – 900 Ages 10 – 15	40 – 1200 15 onwards (lifelong)	Ditto.
Local policy in Paris, France (Abadie, 2000; Burgard et al., 1999) No national guideline.	Universal screening since 1979 Screen ASAP. Treat by day 14 (median)	>600 µmol/ L	120 – 420 Until age 10	<1200 – 1500 After 10 years (Without formula)	French policies are unusually relaxed due to cultural attitudes to food and disability. They previously advised discontinuation of formula at age 5 (see Rey, Abadie, Plainguet & Rey 1996).

In addition to rapid initiation of treatment, guidelines universally recommend stricter *control* (of plasma phe) in childhood (upto ages 10-12). This likewise minimises the developing CNS' exposure to phe and has compelling evidential support. Prospective trials noted severe intellectual impairments following treatment discontinuation/ relaxation before ages five (Cabalska et al., 1977) and seven (Azen et al., 1996) but suffered compliance and randomisation problems as subjects aged, limiting their utility (Poustie & Wildgoose, 2010). However, nationwide prospective cohort studies correlating long-term control data with neurocognitive outcomes (Azen et al., 1991; Smith, Beasley & Ades, 1991) and meta-analyses combining these and other observational data with shorter trials (Waisbren et al., 2007; Fonnesbeck et al., 2012; Albrecht et al., 2009) confirmed that strict control until age ~12 are essential to prevent severe neurocognitive defects resembling those seen in untreated PKU.

3.1: Lifelong treatment is universally recommended.

Whilst guidelines permit relaxation of control in adolescence and adulthood, all recommended lifelong treatment (table 2) as subtle defects occur following cessation. Indeed, meta-analyses suggest control still influences intelligence (Waisbren et al., 2007; Fonnesbeck et al., 2012) and executive functions (Albrecht et al., 2009) from ages 12-18 though the relationships weaken, suggesting control becomes less important. The importance of treatment after age 18 is less clear as lifelong trials are impractical and the retrospective designs most practicable cannot isolate defects from those attributable to control during earlier life (van Wegburg et al., 2017). However, data support lifelong continuation. Trials link even short-term treatment cessation in adulthood to mood and

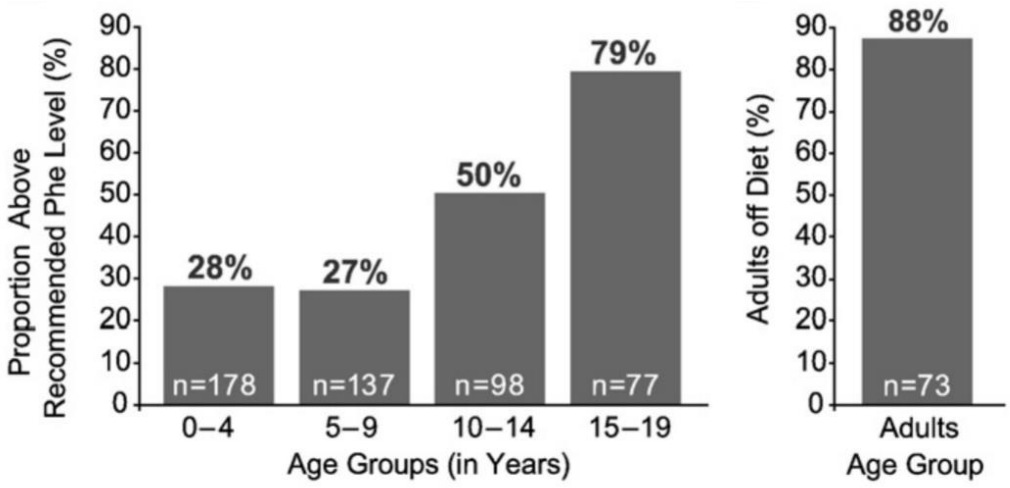
cognitive defects (ten Hoedt et al., 2011) that improve if treatment recommences (Schmidt, Burgard & Rupp., 1996). Whilst their significance is unclear, white matter abnormalities also co-vary with concurrent plasma phe concentrations in adulthood (Anderson & Leuzzi, 2010) and reverse upon treatment recommencement (Cleary et al., 1995). Another rationale for lifelong treatment is uncertainty, as the first recipients of treatment are only ~60 years old.

4.0: Suboptimal control and health remain common despite near-universal treatment.

Treatment from birth is now universally recommended in most developed nations (Therrell et al., 2015; Hagedorn, van Berkel, Hammerschmidt, Lhotáková & Saludes, 2013) and considered successful (Blau et al., 2010; NIH CDP 2001; van Wegburg et al., 2017). Severe disabilities requiring residential care have become rare (Koch & de la Cruz, 1999), early-treated adults typically lead independent lives (Koch et al., 2002) and neurocognitive outcomes generally fall close to or within normal ranges (Waisbren et al., 2007; Fonnesbeck et al., 2012; Beasley Costello & Smith, 1994; Azen et al., 1996). Despite these improvements, average neurocognitive outcomes among early-treated cohorts remain worse than healthy controls and siblings, the most informative comparators (DeRoche & Welsh, 2008; Enns et al., 2010). Neurocognitive outcomes also vary substantially (Waisbren et al., 2007; Fonnesbeck et al., 2012; Albrecht et al., 2009) and in some individuals intelligence remains low enough to affect education and employment prospects (Koch et al., 2002; Beasley Costetllo & Smith, 1994). Given the aforementioned links between control and neurocognitive outcomes, these below-normal neurocognitive outcomes would be expected to coexist with (and be partly attributable to) suboptimal control, which does remain common (figure 3). Few patients maintain it into adolescence, many adults cease treatment

altogether but most concerningly, many children lose control before ages 6-10 when it matters most (MacDonald, Nanuwa, Parkes, Nathan & Chauhan, 2011; Enns et al., 2010).

Figure 3. Percentage of patients maintaining control or treatment at different lifestages



Graph from Enns et al., (2010). Data from Walter et al., (2002).

Whilst below-normal neurocognitive outcomes are most concerning, recent evidence suggests other aspects of health are commonly below-normal (and variable) in PKU including bone health, health related quality of life (HR-QoL), gastrointestinal comfort and adiposity levels (Enns et al., 2010; van Wegberg et al., 2017). This review next considers how limitations of conventional treatment may contribute to these suboptimal health outcomes and how a novel treatment called Glycomacropeptide may help.

4.1: Dietary treatment is burdensome and compliance proves challenging.

Practical treatment information comes from evidence-based guidelines that are not cited throughout (van Wegburg et al., 2017; Smith et al, 1993). Whilst treatment appears simple (regulate phe intake) in practice therapeutic diets are restrictive as phe is ubiquitous in proteins meaning most protein-containing foods are restricted, even staples such as bread

and potatoes (table 3). Only very low-protein foods such as certain fruits, vegetables, fats and phe-free medical foods are consumable in normal quantities.

Table 3. Dietary protein restriction during dietary treatment.

High-protein foods.	**Foods containing aspartame.**	**Moderate-protein foods.**	**Low-protein foods.**
Cannot be consumed.	Cannot be Consumed (metabolised to phe)	Amounts consumable depend on severity/ tolerance. Exchange system expresses tolerance in food terms.	Can be consumed in normal quantities. Excessive consumption may affect plasma phe.
Meat, fish, eggs, most cheeses, nuts, seeds, soya, Quorn, tofu, bread products (containing flour).	Certain fizzy drinks, squashes, cordials, desserts, crisps and chewing gums. Other artificial and natural sweeteners usually acceptable.	Jacket potato 80g* Peas 25g* Baked beans 20g* Corn flakes 15g* Milk 30ml* *These quantities contain 1 exchange, which is 1g Protein & 50mg phe. Patients are assigned X exchanges per day based on their tolerance.	Most fruits and vegetables. Fats such as butter and vegetable oils. Tea, coffee. Pure juice. Rice/ coconut milk. Specially manufactured, low-protein breads, pastas, biscuits, crackers.

Adapted from table in previous assignment and Hallam, (2016).

PKU diets are also complicated. The degree of dietary protein restriction required varies according to individual severity. This reflects residual PAH activity (Blau et al., 2010) though In practice plasma phe monitoring and dietary adjustments are performed to determine the daily protein/ phe intake tolerable without plasma phe leaving safe ranges (van Spronsen et al., 2009). This ***tolerance*** is communicated to patients via an ***exchange*** system (table 3). With protein intake restricted, Phe-free protein substitutes (***formula***) are prescribed to meet outstanding amino acid (AA) requirements. Tolerance and formula requirements change over the lifespan as protein requirements and sensitivity to plasma phe change, further complicating management.

Given the restrictiveness and complexity of therapeutic diets, patients and parents find adherence difficult (MacDonald, Gokmen-Ozel, van Rijn & Burgard, 2010; Walter et al., 2002) particularly with formula, poor adherence with which is noted at all lifestages via self-reports (Schulz & Bremer, 1995; Macleod & Ney, 2010) and more objective measures

(Prince et al., 1997). This is primarily attributed to formula's poor acceptability, particularly the taste, inconvenience and low social acceptability (MacDonald, Harris, Rylance, Asplin & Booth, 1997; Macleod & Ney, 2010) limitations that improved versions (e.g. ready-to-drink pouches) have not totally resolved (Macdonald et al., 2004; 2006; Prince et al., 1997). The burden of conventional treatment, particularly formula, is thus likely to contribute to the persistence of suboptimal control and neurocognitive outcomes.

The persistence of poor control and suboptimal health outcomes despite near-universal treatment and the (apparently related) burden of conventional treatment explain continuing interest in novel treatments (reviewed by van Spronsen & Enns, 2010 and Specola & Chiesa, 2017) including GMP.

4.2: Glycomacropeptide may improve several health outcomes.

GMP is a 64 AA glycophosphopeptide released into bovine whey during cheese production and is the only intact protein naturally lacking phenylalanine (Van Calcar & Ney, 2012). Products manufactured from GMP form alternatives to formula that theoretically address several of its shortcomings and may improve suboptimal outcomes in PKU (see Ney & Etzel, 2017 for review).

4.2.1: Improved protein utilisation.

GMP lacks some essential AAs, necessitating supplementation with some synthetic AAs (Van Calcar et al., 2009). Nevertheless, totally replacing formula with GMP increases the intact protein content of diets markedly (from ~20% to ~70%) (Van Calcar & Ney 2012). This would be expected to improve protein utilisation as intact proteins are digested less rapidly and their AA reach plasma more slowly (Gropper & Acosta, 1991) decreasing oxidation (wastage)

and increasing protein utlisation/ synthesis (Metges et al., 2000). Findings of one short trial (showing decreased blood urea nitrogen, indicative of decreased ureagenesis) suggest protein utilisation improves as expected when replacing formula with GMP, though nitrogen balance and protein synthesis were not monitored (Van Calcar et al., 2009). Improved maintenance of bodily proteins, growth and adaptations to exercise may be expected to ensue over longer timeframes (Millward, Layman, Tomé, & Schaafsma, 2008) but are yet to be demonstrated. Improved control may also ensue as improved protein synthesis may incorporate more plasma phe into tissue proteins (the mechanism thought to explain how better adherence to formula improves control: see Macdonald et al., 2005). The effect of GMP upon control is complicated however and discussed later.

4.2.2: Satiety and excess adiposity.

Persistent hunger is reported in PKU (MacLeod, Clayton, van Calcar & Ney, 2010) and obesity is common among females (Gokmen-Ozel et al., 2014). The causation of Obesity is complex (Rocha, Macdonald & Trefz, 2013) but the lack of intact protein in therapeutic diets may contribute as protein is the most satiating macronutrient (Greco et al., 2017) and even 5% increases in intake improve long-term weight maintenance, partly as protein's satiety displaces other macronutrients (Santesso et al., 2012; Clifton, Keogh & Condo, 2014). One short-term trial suggests GMP recreates the satiety of intact proteins better than formula (MacLeod et al, 2010) as would be expected. Therein, a GMP breakfast significantly increased subjective satiety, total plasma AA concentration and ghrelin suppression over formula. The metabolic changes induced are consistent with theoretical explanations for differences in satiety-promotion between nutritional factors (Greco et al., 2017). However, unblinding may have biased subjective satiety ratings as GMP tastes different. Moreover,

reduced food intake is yet to demonstrated over longer timeframes and may not occur (Poppitt et al., 2013) as appetite regulation is complex (Begg & Woods, 2013) and compensatory eating can occur (Benelam, 2009).

4.2.3: Gastrointestinal symptoms.

Formulas are hyperosmolar and may cause gastro-intestinal discomfort by drawing water into the digestive tract (van Wegberg et al., 2017) as non-digestible carbohydrates and enteric feeds can (Clausen, Jorgensen & Mortensen, 1998; Williams, 2008). GMP-based foods have lower osmolalities and may ameliorate symptoms (Ney & Etzel, 2017). However, only uncontrolled studies (Macdonald et al., 1997) and qualitative reports (Ney et al., 2016) have linked symptoms to formula and improvements to GMP. Trials controlling for confounders including the type and dose of formula, other dietary constituents and innate susceptibility to gastrointestinal symptoms are required. Nevertheless, improved gastrointestinal comfort appears plausible.

4.2.3: Bone health.

Meta-analyses suggest the risk of impaired bone mineral density (BMD) is increased in PKU (Hansen & Ney 2014; Enns et al., 2010) though the strongest suggests only ~10% of cases reach clinical significance (Demirdas et al., 2015). Bone health in PKU is complex and poorly understood as most human studies are cross-sectional because formula cannot be ethically withheld for long periods during childhood. A study withholding formula in PKU mice (post-weaning) suggests impairments are partly inherent to PKU but exacerbated by formula and partly rectified by GMP (Solverson, Murali, Litscher, Blank & Ney, 2012). Proposed mechanisms include improved synthesis of collagen, a protein constituent of bone (van

Wegberg et al, 2017) and the higher pH of GMP reducing skeletal buffering, which may resorb bone minerals to maintain acid-based homeostasis (Lemann, Bushinsky & Hamm, 2003). However, BMD improvements are yet to be demonstrated in humans.

4.2.4: Acceptability.

The most consistently reported advantage of GMP over formula is greater acceptability, suggesting GMP addresses a major shortcoming of formula. Superior food making properties (e.g. heat stability) enable the production of a variety of more food-like products (Van Calcar & Ney, 2012). Resultantly, most patients consider GMP foods more palatable (Ney et al., 2016; Lim, van Calcar, Nelson, Gleason & Ney, 2007; Van Calcar et al., 2009; Zaki et al., 2016) and socially acceptable (Ney et al., 2016) than formula. It should be stressed that the strongest of these studies, a randomised crossover trial (Ney et al., 2016) selectively recruited patients finding GMP acceptable, potentially biasing results. Nevertheless, the weight of evidence suggests most subjects prefer GMP-based diets. Greater acceptability may improve HR-QOL, something formulas poor taste impacts upon (Bosch et al., 2015) but HR-QoL tools sensitive to PKU-specific difficulties have been validated only recently (Regnault et al., 2015) and trials are yet to utilise them. Greater acceptability may also indirectly improve control by promoting compliance with therapeutic diets, though the impact of GMP upon control is discussed shortly.

GMP products are yet to elicit improvements in any long-term health outcome in a controlled trial. However, this is partly relates to constraints upon the types of study practicable. The low prevalence of PKU makes attracting funding and recruitment difficult, particularly for long-term trials, formula cannot ethically be withheld during childhood and GMP foods were only recently applied to PKU (~2009). Whilst clearly requiring confirmation,

each of the proposed advantages is theoretically plausible, supported by preliminary data and has the potential to improve therapeutic diets and reduce the burden of PKU.

5.0: The effect of GMP on control is poorly understood, raising concern.

Given the importance of control, GMPs impact upon it is paramount. Whilst improved control may address suboptimal neurocognitive outcomes, equivalent control to formula would constitute a positive finding given GMP's other advantages. However, GMPs effect on control is complicated as, whilst GMP is naturally phe-free, commercially viable purification processes leave GMP contaminated with *residual phe* from other whey constituents (Ney & Etzel 2017; La Clair, Ney, Macleod & Etzel 2009). Commercial GMP products contain ~1.8 mg phe per gram protein (Ney et al., 2016; Vitaflo*, 2017*) raising safety concerns, particularly among children that tolerate less phe. As the most recent evidence-based guidelines acknowledge (van Wegberg, 2017) these concerns remain unresolved.

5.1: Effect on control in adults (over 16).

In adults, every study this non-systematic review located administered GMP identically. GMP totally replaced formula then the additional phe introduced was compensated for by reducing exchanges so diets contained equal phe (Van Calcar & Ney, 2012). A mouse study (Ney, Hull, van Calcar, Liu & Etzel, 2007) and case study (Ney et al., 2009) suggest GMP improves control over formula by ~10% (average) despite diets containing equal phe. Improvements are attributed to three mechanisms. GMP naturally contains more threonine and isoleucine than other proteins (Ney et al., 2009) and these LNAA may decrease intestinal phe absorption by competitively inhibiting LAT-1 transporters as LNAA treatment

does (Sanjuro et al., 2003; Matalon et al., 2006; 2007). The formulation of GMP used in these studies and others by Ney & colleagues (Glytactin) is also supplemented to contain higher quantities of other LNAA (otherwise limiting in GMP) than formula (150% of USA RDI for Tyrosine, 130% for other limiting LNAA). Secondly, GMPs greater acceptability may improve compliance in free-living subjects. Compliance with protein substitutes involves consuming sufficient quantities (Macdonald et al., 2005; Duran et al., 1999) and distributing 'doses' over 3-4 meals, which both promote control by increasing protein utilisation to move more plasma phe into tissue proteins (Mönch, Herrmann, Brösicke, Schöffer & Keller, 1996; MacDonald et al., 1996). Compliance with the wider therapeutic diet prevents the consumption of excess phe-containing foods and inadequate energy, which causes catabolism and mobilises phe from tissue proteins. The case study participant reported greater compliance with the GMP diet, citing its acceptability (Ney et al., 2009). Finally, GMP may improve protein utilisation over formula, incorporating more plasma phe into tissue proteins (section 3.2.1). Ensuing control improvements may improve cognitive outcomes or increase tolerance, permitting dietary relaxation.

Control improvements have not been recreated in controlled trials, however. A self-controlled inpatient trial (van Calcar et al., 2009) and a randomised, crossover involving outpatients (Ney et al., 2016) both reported no significant difference in control after GMP replaced formula. The discrepancy in findings remains unexplained. It may relate to the mouse model and single subject employed in earlier studies poorly representing PKU populations or bias in the earlier, weaker studies. Bias is minimised by design features like randomisation which tend to cluster in stronger studies (Deeks et al., 2003) and generally (though not always) overstates an interventions benefit (Higgins, Altman & Sterne, 2008).

Heterogenity may also contribute. For example, the improvements in compliance and control noted among free-living subject(s) by Ney et al., (2009) but not Ney et al., (2016) may relate to the former study comparing ready-to-eat, food-like GMP products to powdered formulas requiring weighing and preparation, whilst the later study mainly compared GMP to ready-to-drink liquid formulas that form a stronger control as they are more acceptable (Macdonald et al., 2004; 2006). Different plasma phe assessment methods may similarly influence results (Gregory, Yu & Singh 2007). Nevertheless, it must be stressed that the broadly equivalent control demonstrated is a positive finding given GMPs other advantages.

Additional studies are required to confirm this as only two controlled human trials with limitations were located. One administered GMP for only 4 days, lacked randomisation and used an inpatient setting lacking ecological validity (van Calcar et al., 2009). The strongest (Ney et al., 2016) purposively recruited patients finding GMP acceptable, potentially biasing results. It also demonstrated maintenance of control despite the GMP condition providing 88 ± 6mg/ day additional phe as dietary exchanges were not reduced to compensate in the GMP condition. This may be interpreted as suggesting 'control would be equivalent if patients complied'. However, difficulties removing exchanges are concerning and may compromise control in some individuals. In addition, every GMP-centric review (Van Calcar & Ney, 2012; Ney, Blank & Hansen, 2014; Ney & Etzel, 2017) and primary study previously cited are from one group with commercial interests in GMP products (see conflicts of interest in Ney et al., 2016) raising objectivity concerns (Oxman & Guyatt, 1993). These reviews have not specifically investigated GMP's effect on control and are narrative so have not synthesized or evaluated the entire evidence base using transparent, systematic

methods such as those of the Cochrane collaboration (Green et al., 2011) and PRISMA (Liberati et al., 2009) which evaluate the contribution of bias (Mulrow, 1994; Oxman & Guyatt, 1993) to results. There is therefore need for such a review, to clarify the effect of GMP upon control in adults (over 16). It should also clarify GMPs effect upon compliance, particularly whether GMP promotes compliance and whether non-compliance may compromise control in some cases.

5.2: Effect on control in children (3 - 16).

GMP foods contain excessive phe and insufficient energy for children under three (van Calcar et al., 2012). However, several have approval for older children in England (Cambrooke, 2018; Vitaflo, 2017) despite their effect upon control remaining controversial. Children require more protein (per kg BM) than adults for growth (WHO, 2007) but tolerate less phe (table 6) so formula constitutes more of their diets (van Calcar et al., 2012). The substitution method used in adults (100% replacement) appears unsuitable as the phe introduced can be excessive given the lower tolerances and that insufficient exchanges are consumed to remove in compensation. Studies involving children have instead investigated partially replacing formula with GMP, though only three were located. In one, control was maintained on average when comparing 50% GMP & 50% formula to 100% formula over nine weeks (Zaki et al., 2016). Another reported improved control when replacing formula with GMP for 3 days in infants, children and adults (Abdel-Salam & Effat, 2010). However, both lack randomisation and pilot-test incomplete GMP formulations differing from commercial products in their organoleptics, LNAA content and phe content. Both manuscripts also omit important details such as compliance data and the type of formula used.

Whilst Zaki et al., (2016) reported equivalent control after replacing the same proportion of formula with GMP in each individual (50%) another study used a different approach. Individualised titrations (adjusting the amount of formula replaced by GMP) were necessary to maintain control during a six month study (Daly, Evans, Chahal, Santra & MacDonald, 2017) in which ethics precluded control loss. The average replacement possible was 50% (agreeing with Zaki and colleagues) though the amount varied from 30 to 100%. This suggests maintenance of control depends on individualised adjustments though the design was similarly limited. Again a preliminary form of GMP was used though a self-controlled design was not. These are standard in PKU as between-subject comparisons of plasma phe are problematic since tolerance, intensity of treatment and baseline control vary. Comparisons with retrospective control data collected under non-experimental conditions were thus necessary, potentially introducing confounding.

Overall, the effect of GMP on control in children is less well understood than in adults despite GMP being licenced for use in this group and control being crucial. This informal review noted only three pilot studies using heterogeneous, bias-prone designs, poor reporting and preliminary GMP formulations. These suggest maintenance of control in children is possible following GMP substitution provided individual-level adjustments are performed but the suggestion requires confirmation, ideally through randomised, controlled longitudinal trials with self-controlled designs, that perform patient-level adjustments if required. A systematic review is required to clarify GMPs impact on control in children. It should use methods mirroring those described in adults to synthesize the entire evidence base then consider the contribution of design quality, bias, heterogeneity and compliance to results. Ideally, the review should describe inter-individual variation in phe responses (of the

amount of formula safely replaceable with GMP) and identify their determinants, as this may help to direct treatment.

6.0: Conclusion.

Before treatment was discovered PKU was associated with severe neurocognitive deficits requiring residential care. Treatment restricts dietary phenylalanine to control plasma phe, preventing its accumulation to toxic levels. As the developing CNS is most vulnerable to phe, treatment is initiated at birth and stricter control is essential during childhood but lifelong adherence is universally recommended as defects occur with later cessation. Since the widespread adoption of dietary treatment the severe disabilities previously seen have become rare but intelligence and other aspects of health remain below-normal even among early-treated patients, partly as dietary treatment has limitations. Adherence proves difficult as therapeutic diets are restrictive, complicated and formulas lack acceptability. GMP is a novel alternative to formula that emerging evidence suggests has the potential to improve therapeutic diets and patients health in several ways. However, GMP contains some phenylalanine raising concerns around its impact on control, particularly among children. Despite GMP having UK approval for over 3s, these concerns remain unresolved and the recent European evidence-based guidelines offer no advice regarding GMP use.

The evidence reviewed herein suggests adults generally maintain control after GMP replaces formula, a supportive finding given GMPs many advantages. This requires confirmation however, as only two controlled trials were located. One was unrealistically short, the other suffered dietary compliance problems that may impair control in some individuals and both had heterogeneous designs. Existing reviews are similarly limited and come from one group commercially linked to products, raising objectivity concerns. None specifically consider

control and all are narrative, suggesting bias may affect conclusions. A systematic review is thus necessary to clarify the effect of GMP on control in adults by synthesizing and evaluating the entire evidence base using pre-determined methods to consider the contribution of design quality, bias, heterogeneity and compliance levels to results. It should also clarify the utility partially replacing formula with GMP in adults.

GMPs effect on control in children is less clear, which is concerning given the importance of childhood control. Only three, small preliminary studies were located, most with poor reporting and all with bias-prone, heterogeneous designs that administered preliminary GMP formulations unrepresentative of those used clinically. Taken together, they suggest maintenance of control is possible following partial replacement of formula with GMP in children but the proportion of formula replicable without control loss is likely to vary individually, necessitating individualised monitoring and adjustments. However given the few, limited studies reviewed, a systematic review mirroring that described for adults is necessary to clarify matters. It should ideally attempt to describe and explore inter-individual variation in responses to inform practice (e.g. by helping to direct treatment).

7.0: References.

Abadie, V. (2000). PKU treatment in France. *Presentations to the 2000 NIH Consensus Development Conference on Phenylketonuria (PKU): Screening and Management*. Retrieved from: https://www.nichd.nih.gov/publications/pubs/pku/Pages/sub20.aspx

Abdel-Salam, A. M., & Effat, L. K. (2010). Preparation and evaluation of a novel therapeutic dairy-based drink for phenylketonuria. *North American Journal of Medical Sciences*, 2(2), 66.

Albrecht, J., Garbade, S. F., & Burgard, P. (2009). Neuropsychological speed tests and blood phenylalanine levels in patients with phenylketonuria: A meta-analysis. *Neuroscience & Biobehavioral Reviews*, *33*(3), 414–421.

Anderson, P. J., & Leuzzi, V. (2010). White matter pathology in phenylketonuria. *Molecular Genetics and Metabolism*, *99*, S3–S9.

Azen, C. G., Koch, R., Friedman, E. G., Berlow, S., Coldwell, J., Krause, W., … Peterson, R. (1991). Intellectual development in 12-year-old children treated for phenylketonuria. *American Journal of Diseases of Children (1960)*, *145*(1), 35–39.

Azen, C., Koch, R., Friedman, E., Wenz, E., & Fishier, K. (1996). Summary of findings from the United States Collaborative Study of children treated for phenylketonuria. *European Journal of Pediatrics*, 155, S29–S32.

Beasley, M. G., Costello, P. M., & Smith, I. (1994). Outcome of treatment in young adults with phenylketonuria detected by routine neonatal screening between 1964 and 1971. *QJM: An International Journal of Medicine*, 87(3), 155–160.

Begg, D. P., & Woods, S. C. (2013). The endocrinology of food intake. *Nature Reviews Endocrinology*, 9(10), 584–597.

Benelam, B. (2009). Satiation, satiety and their effects on eating behaviour. *Nutrition Bulletin*, 34(2), 126–173.

Bickel, H. (1996). The first treatment of phenylketonuria. *European Journal of Pediatrics*, 155(S1), S2–S3.

Bickel, H., Gerrard, J., & Hickmans, E. M. (1954). The Influence of Phenylalanine Intake on the Chemistry and Behaviour of a Phenylketonuria Child. *Acta Paediatrica*, 43(1), 64–77.

Blau, N., Hennermann, J. B., Langenbeck, U., & Lichter-Konecki, U. (2011). Diagnosis, classification, and genetics of phenylketonuria and tetrahydrobiopterin (BH4) deficiencies. *Molecular Genetics and Metabolism*, *104*, S2–S9.

Blau, N., van Spronsen, F. J., & Levy, H. L. (2010). Phenylketonuria. *The Lancet*, *376*(9750), 1417–1427.

Bosch, A. M., Burlina, A., Cunningham, A., Bettiol, E., Moreau-Stucker, F., Koledova, E., … Regnault, A. (2015). Assessment of the impact of phenylketonuria and its treatment on quality of life of patients and parents from seven European countries. *Orphanet Journal of Rare Diseases*, 10(1), 80

Burgard, P., Bremer, H. J., Bührdel, P., Clemens, P. C., Mönch, E., Przyrembel, H., … Ullrich, K. (1999). Rationale for the German recommendations for phenylalanine level control in phenylketonuria 1997. *European Journal of Pediatrics*, 158(1), 46–54.

Cabalska, B., Duczyńska, N., Borzymowska, J., Zorska, K., Koślacz-Folga, A., & Bożkowa, K. (1977). Termination of dietary treatment in phenylketonuria. *European Journal of Pediatrics*, 126(4), 253–262.

Cambrooke (2018). Cambrooke therapeutics UK: Glytactin BetterMilk information page and data card. Retrieved from: http://cambrooketherapeutics.co.uk/glytactin-bettermilk-15/

Clausen, M. R., Jørgensen, J., & Mortensen, P. B. (1998). Comparison of diarrhoea induced by ingestion of fructooligosaccharide Idolax and disaccharide lactulose: role of osmolarity versus fermentation of malabsorbed carbohydrate. *Digestive Diseases and Sciences*, 43(12), 2696–2707.

Cleary, M. A., Walter, J. H., Wraith, J. E., White, F., Tyler, K., & Jenkins, J. P. (1995). Magnetic resonance imaging in phenylketonuria: reversal of cerebral white matter change. *The Journal of Pediatrics*, *127*(2), 251–255.

Clifton, P. M., Condo, D., & Keogh, J. B. (2014). Long term weight maintenance after advice to consume low carbohydrate, higher protein diets – A systematic review and meta analysis. *Nutrition, Metabolism and Cardiovascular Diseases*, 24(3), 224–235.

Daly, A., Evans, S., Chahal, S., Santra, S., & MacDonald, A. (2017). Glycomacropeptide in children with phenylketonuria: does its phenylalanine content affect blood phenylalanine control? *Journal of Human Nutrition and Dietetics*, 30, 515–523.

Deeks, J. J., Dinnes, J., D'amico, R., Sowden, A. J., Sakarovitch, C., Song, F., ... & Altman, D. G. (2003). Evaluating non-randomised intervention studies. *Health technology assessment* (Winchester, England), 7(27), iii-x. 1-173

de Groot, M. J., Hoeksma, M., Blau, N., Reijngoud, D. J., & van Spronsen, F. J. (2010). Pathogenesis of cognitive dysfunction in phenylketonuria: Review of hypotheses. *Molecular Genetics and Metabolism*, *99*, S86–S89.

Demirdas, S., Coakley, K. E., Bisschop, P. H., Hollak, C. E. M., Bosch, A. M., & Singh, R. H. (2015). Bone health in phenylketonuria: a systematic review and meta-analysis. *Orphanet Journal of Rare Diseases*, 10(1), 17.

DeRoche, K., & Welsh, M. (2008). Twenty-Five Years of Research on Neurocognitive Outcomes in Early-Treated Phenylketonuria: Intelligence and Executive Function. *Developmental Neuropsychology*, 33(4), 474–504.

Duran, G. P., Rohr, F. J., Slonim, A., Guttler, F., & Levy, H. L. (1999). Necessity of Complete Intake of Phenylalanine-free Amino Acid Mixture for Metabolic Control of Phenylketonuria. *Journal of the American Dietetic Association*, 99, 1559–1563.

Enns, G. M., Koch, R., Brumm, V., Blakely, E., Suter, R., & Jurecki, E. (2010). Suboptimal outcomes in patients with PKU treated early with diet alone: Revisiting the evidence. *Molecular Genetics and Metabolism*, *101*(2–3), 99–109.

Fonnesbeck, C. J., McPheeters, M. L., Krishnaswami, S., Lindegren, M. L., & Reimschisel, T. (2013). Estimating the probability of IQ impairment from blood phenylalanine for phenylketonuria patients: a hierarchical meta-analysis. *Journal of Inherited Metabolic Disease*, *36*(5), 757–766.

Gokmen Ozel, H., Ahring, K., Bélanger-Quintana, A., Dokoupil, K., Lammardo, A. M., Robert, M., … MacDonald, A. (2014). Overweight and obesity in PKU: The results from 8 centres in Europe and Turkey. *Molecular Genetics and Metabolism Reports*, 1, 483–486.

Greco, E., Winquist, A., Lee, T. J., Collins, S., & Lebovic, Z. (2017). The Role of Source of Protein in Regulation of Food Intake, Satiety, Body Weight and Body Composition. *Journal of Nutrition Health and Food Engineering*, 6(6), 186-193.

Green, S., Higgins, J. P. T., Alderson P., Clarke M., Mulrow C. D., & Oxman A. D. (2011) Chapter 1: Introduction. In: Higgins JPT, Green S (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0.* The Cochrane Collaboration, 2011. Available from www.handbook.cochrane.org.

Gregory, C. O., Yu, C., & Singh, R. H. (2007). Blood phenylalanine monitoring for dietary compliance among patients with phenylketonuria: comparison of methods. *Genetics in Medicine*, 9(11), 761.

Gropper, S. S., & Acosta, P. B. (1991). Effect of simultaneous ingestion of L-amino acids and whole protein on plasma amino acid and urea nitrogen concentrations in humans. *Journal of Parenteral and Enteral Nutrition*, 15(1), 48–53.

Guthrie, R., & Susi, A. (1963). A simple phenylalanine method for detecting phenylketonuria in large populations of newborn infants. *Pediatrics*, *32*, 338–343.

Hagedorn, T. S., van Berkel, P., Hammerschmidt, G., Lhotáková, M., & Saludes, R. P. (2013). Requirements for a minimum standard of care for phenylketonuria: the patients' perspective. *Orphanet Journal of Rare Disease*, 8(1), 191-198.

Hallam., P (2016). Dietary Information for the treatment of phenylketonuria: The National society for Phenylketonuria. Retrieved from: http://www.nspku.org/publications/publication/dietary-information-treatment-phenylketonuria-20162017

Hansen, K. E., & Ney, D. (2014). A systematic review of bone mineral density and fractures in phenylketonuria. *Journal of Inherited Metabolic Disease*, 37(6), 875–880.

Hardelid, P., Cortina-Borja, M., Munro, A., Jones, H., Cleary, M., Champion, M. P., … Dezateux, C. (2007). The Birth Prevalence of PKU in Populations of European, South Asian and Sub-Saharan African Ancestry Living in South East England. *Annals of Human Genetics*, *0*(0)

Higgins, J. P. T., Altman, D. G., & Sterne J. A. C., (2017) Chapter 8: Assessing risk of bias in included studies. In Higgins, J. P. T., Churchill, R., Chandler, J., Cumpston, M. S. (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions* version 5.2.0, updated June 2017 (pp. 8:1-8:73) Cochrane

Jervis, G. A. (1937). Phenylpyruvic Oligophrenia: Introductory study of fifty cases of mental deficiency associated with excretion of phenylpyruvic acid. *Archives of Neurology & Psychiatry*, *38*(5), 944-963.

Koch, R., Burton, B., Hoganson, G., Peterson, R., Rhead, W., Rouse, B., … Guttler, F. (2002). Phenylketonuria in adulthood: a collaborative study. *Journal of Inherited Metabolic Disease*, 25(5), 333–346.

Koch, R., & de la Cruz, F. (1999). Historical aspects and overview of research on phenylketonuria. *Mental Retardation and Developmental Disabilities Research Reviews*, *5*(2), 101–103.

LaClair, C. E., Ney, D. M., MacLeod, E. L., & Etzel, M. R. (2009). Purification and Use of Glycomacropeptide for Nutritional Management of Phenylketonuria. *Journal of Food Science*, 74(4), E199–E206.

Lemann, J., Bushinsky, D. A., & Hamm, L. L. (2003). Bone buffering of acid and base in humans. American Journal of Physiology-Renal Physiology, 285(5), F811–F832.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., … Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Medicine, 6(7), e1000100.

Lim, K., van Calcar, S. C., Nelson, K. L., Gleason, S. T., & Ney, D. M. (2007). Acceptable low-phenylalanine foods and beverages can be made with glycomacropeptide from cheese whey for individuals with PKU. *Molecular Genetics and Metabolism*, 92(1), 176–178.

Loeber, J. G. (2007). Neonatal screening in Europe; the situation in 2004. *Journal of Inherited Metabolic Disease*, *30*(4), 430–438.

Longo, N. (2009). Disorders of biopterin metabolism. *Journal of Inherited Metabolic Disease*, *32*(3), 333–342.

Lord, J., Thomason, M. J., Littlejohns, P., Chalmers, R. A., Bain, M. D., Addison, G. M., … Seymour, C. A. (1999). Secondary analysis of economic data: a review of cost-benefit studies of neonatal screening for phenylketonuria. *Journal of Epidemiology & Community Health*, 53(3), 179–186.

MacDonald, A. (2005). Protein substitute dosage in PKU: how much do young patients need? *Archives of Disease in Childhood*, 91(7), 588–593.

MacDonald, A., Davies, P., Evans, S., Daly, A., Hall, S. K., Hendriksz, C., ... Lee, P. (2006). 'Ready to drink' protein substitute is easier is for people with phenylketonuria. *Journal of inherited metabolic disease*, 29(4), 526-531.

MacDonald, A., Gokmen-Ozel, H., van Rijn, M., & Burgard, P. (2010). The reality of dietary compliance in the management of phenylketonuria. *Journal of Inherited Metabolic Disease*, 33(6), 665–670.

MacDonald, A., Harris, G., Rylance, G., Asplin, D., & Booth, I. W. (1997). Abnormal feeding behaviours in phenylketonuria. *Journal of Human Nutrition and Dietetics*, 10(3), 163–170.

MacDonald, A., Lilburn, M., Cochrane, B., Davies, P., Daly, A., Asplin, D., … Robinson, P. (2004). A new, low-volume protein substitute for teenagers and adults with phenylketonuria. *Journal of Inherited Metabolic Disease*, 27(2), 127–135.

MacDonald, A., Nanuwa, K., Parkes, L., Nathan, M., & Chauhan, D. (2011). Retrospective, observational data collection of the treatment of phenylketonuria in the UK, and associated clinical and health outcomes. *Current Medical Research and Opinion*, 27(6), 1211–1222.

MacDonald, A., Rylance, G., Hall, S. K., Asplin, D., & Booth, I. W. (1996). Factors affecting the variation in plasma phenylalanine in patients with phenylketonuria on diet. *Archives of Disease in Childhood*, 74(5), 412–417.

MacLeod, E. L., Clayton, M. K., van Calcar, S. C., & Ney, D. M. (2010). Breakfast with glycomacropeptide compared with amino acids suppresses plasma ghrelin levels in individuals with phenylketonuria. *Molecular Genetics and Metabolism*, 100(4), 303–308.

MacLeod, E. L., & Ney, D. M. (2010). Nutritional Management of Phenylketonuria. *Annales Nestlé* (English Ed.), 68(2), 58–69.

Matalon, R., Michals-Matalon, K., Bhatia, G., Burlina, A. B., Burlina, A. P., Braga, C., … Novikov, P. (2007). Double blind placebo control trial of large neutral amino acids in treatment of PKU: effect on blood phenylalanine. *Journal of Inherited Metabolic Disease*, 30(2), 153–158.

Matalon, R., Michals-Matalon, K., Bhatia, G., Grechanina, E., Novikov, P., McDonald, J. D., … Guttler, F. (2006). Large neutral amino acids in the treatment of phenylketonuria (PKU). *Journal of Inherited Metabolic Disease*, 29(6), 732–738.

Mazur, A., Jarochowicz, S., Ołtarzewski, M., Sykut-Cegielska, J., Gradowska, W., Januszek-Trzciąkowska, A., … Kwolek, A. (2011). An Investigation of the Neurological and Neuropsychiatric Disturbances in Adults with Undiagnosed and/or Untreated Phenylketonuria in Poland: Journal of Applied Research in Intellectual Disabilities. *Journal of Applied Research in Intellectual Disabilities*, *24*(5), 482–488

Metges, C. C., El-Khoury, A. E., Selvaraj, A. B., Tsay, R. H., Atkinson, A., Regan, M. M., … Young, V. R. (2000). Kinetics of L-[1-13C] leucine when ingested with free amino acids, unlabelled or intrinsically labelled casein. *American Journal of Physiology-Endocrinology And Metabolism*, 278(6), E1000–E1009.

Millward, D. J., Layman, D. K., Tomé, D., & Schaafsma, G. (2008). Protein quality assessment: impact of expanding understanding of protein and amino acid needs for optimal health–. *The American Journal of Clinical Nutrition*, 87(5), 1576S–1581S.

Mönch, E., Herrmann, M. E., Brösicke, H., Schöffer, A., & Keller, M. (1996). Utilisation of amino acid mixtures in adolescents with phenylketonuria. *European Journal of Pediatrics*, 155(1), S115–S120.

MRC (1963). Treatment of phenylketonuria: Report to the Medical Research Council (MRC) of the conference on phenylketonuria. *British Medical Journal*, 1(5347), 1691–1697.

Mulrow, C. D. (1994). Systematic Reviews: Rationale for systematic reviews. *British Medical Journal*, 309(6954), 597–599.

Murphy, G. H., Johnson, S. M., Amos, A., Weetch, E., Hoskin, R., Fitzgerald, B., … Lee, P. (2008). Adults with untreated phenylketonuria: out of sight, out of mind. *The British Journal of Psychiatry*, *193*(6), 501–502.

NIH CDP. (2001). National Institutes of Health Consensus Development Panel. Conference Statement: phenylketonuria: screening and management, October 16-18, 2000. *Pediatrics*, 108(4), 972–982.

Ney, D. M., Blank, R. D., & Hansen, K. E. (2013). Advances in the nutritional and pharmacological management of phenylketonuria. *Current Opinion in Clinical Nutrition and Metabolic Care*, 17(1), 61-8.

Ney, D. M., & Etzel, M. R. (2017). Designing medical foods for inherited metabolic disorders: why intact protein is superior to amino acids. *Current Opinion in Biotechnology*, 44, 39–45.

Ney, D. M., Gleason, S. T., van Calcar, S. C., MacLeod, E. L., Nelson, K. L., Etzel, M. R., … Wolff, J. A. (2009). Nutritional management of PKU with glycomacropeptide from cheese whey. *Journal of Inherited Metabolic Disease*, 32(1), 32–39.

Ney, D. M., Hull, A. K., van Calcar, S. C., Liu, X., & Etzel, M. R. (2008). Dietary glycomacropeptide supports growth and reduces the concentrations of phenylalanine in plasma and brain in a murine model of phenylketonuria. *The Journal of Nutrition*, 138(2), 316–322.

Ney, D. M., Stroup, B. M., Clayton, M. K., Murali, S. G., Rice, G. M., Rohr, F., & Levy, H. L. (2016). Glycomacropeptide for nutritional management of phenylketonuria: a randomized, controlled, crossover trial. *The American Journal of Clinical Nutrition*, 104(2), 334–345.

Paine, R. S. (1957). The variability in manifestations of untreated patients with phenylketonuria (phenylpyruvic aciduria). *Pediatrics*, *20*(2), 290–302.

Poppitt, S. D., Strik, C. M., McArdle, B. H., McGill, A.-T., & Hall, R. S. (2013). Evidence of Enhanced Serum Amino Acid Profile but Not Appetite Suppression by Dietary Glycomacropeptide (GMP): A Comparison of Dairy Whey Proteins. *Journal of the American College of Nutrition*, 32(3), 177–186.

Poustie, V. J., & Wildgoose, J. (2010). Dietary interventions for phenylketonuria. *Cochrane Database of Systematic Reviews*.

Prince, A. P., McMurry, M. P., & Buist, N. R. M. (1997). Treatment products and approaches for phenylketonuria: improved palatability and flexibility demonstrate safety, efficacy and acceptance in US clinical trials. *Journal of Inherited Metabolic Disease*, 20(4), 486–498.

Public Health England (2016) Diabetes Prevalence Model: Analysis of estimates of diabetes prevalence across England. April 2016. Retrieved from: https://www.gov.uk/government/publications/diabetes-prevalence-estimates-for-local-populations

Pietz, J., Kreis, R., Rupp, A., Mayatepek, E., Rating, D., Boesch, C., & Bremer, H. J. (1999). Large neutral amino acids block phenylalanine transport into brain tissue in patients with phenylketonuria. *Journal of Clinical Investigation*, *103*(8), 1169.

Regnault, A., Burlina, A., Cunningham, A., Bettiol, E., Moreau-Stucker, F., Benmedjahed, K., & Bosch, A. M. (2015). Development and psychometric validation of measures to assess the impact of phenylketonuria and its dietary treatment on patients' and parents' quality of life: the phenylketonuria – quality of life (PKU-QOL) questionnaires. *Orphanet Journal of Rare Diseases*, 10(1).

Rey, F., Abadie, V., Plainguet, F., & Rey, J. (1996). Long-term follow up of patients with classical phenylketonuria after diet relaxation at 5 years of age. The Paris Study. *European Journal of Pediatrics*, 155 Suppl 1, S39-44.

Rocha, J. C., MacDonald, A., & Trefz, F. (2013). Is overweight an issue in phenylketonuria? *Molecular Genetics and Metabolism*, 110, S18–S24.

Sanjurjo, P., Aldamiz, L., Georgi, G., Jelinek, J., Ruiz, J. I., & Boehm, G. (2003). Dietary threonine reduces plasma phenylalanine levels in patients with hyperphenylalaninemia. *Journal of Pediatric Gastroenterology and Nutrition*, 36(1), 23–26.

Santesso, N., Akl, E. A., Bianchi, M., Mente, A., Mustafa, R., Heels-Ansdell, D., & Schünemann, H. J. (2012). Effects of higher-versus lower-protein diets on health outcomes: a systematic review and meta-analysis. *European Journal of Clinical Nutrition*, 66(7), 780.

Schmidt, E., Burgard, P., & Rupp, A. (1996). Effects of concurrent phenylalanine levels on sustained attention and calculation speed in patients treated early for phenylketonuria. *European Journal of Pediatrics*, 155, S82–S86.

Schulz, B., & Bremer, H. J. (1995). Nutrient intake and food consumption of adolescents and young adults with phenylketonuria. *Acta Paediatrica* (Oslo, Norway: 1992), 84(7), 743–748.

Smith, I., Beasley, M. G., & Ades, A. E. (1990). Intelligence and quality of dietary treatment in phenylketonuria. *Archives of Disease in Childhood*, *65*(5), 472–478.

Smith, I., Beasley, M. G., & Ades, A. E. (1991). Effect on intelligence of relaxing the low phenylalanine diet in phenylketonuria. *Archives of Disease in Childhood*, 66(3), 311–316.

Smith, I., Cockburn, F., Barwell, B., Brenton, D., Chapple, J., Clark, B., … Laing, S. (1993). Recommendations on the dietary management of phenylketonuria. *Archives of Disease in Childhood*, 68(3), 426–427.

Solverson, P., Murali, S. G., Litscher, S. J., Blank, R. D., & Ney, D. M. (2012). Low Bone Strength Is a Manifestation of Phenylketonuria in Mice and Is Attenuated by a Glycomacropeptide Diet. *PLoS ONE*, 7(9), e45165.

Spécola, N., & Chiesa, A. (2017). Alternative Therapies for PKU. *Journal of Inborn Errors of Metabolism and Screening*, (5) 1-5.

Therrell, B. L., Padilla, C. D., Loeber, J. G., Kneisser, I., Saadallah, A., Borrajo, G. J. C., & Adams, J. (2015). Current status of newborn screening worldwide: 2015. *Seminars in Perinatology*, 39(3), 171–187.

van Calcar, S. C., MacLeod, E. L., Gleason, S. T., Etzel, M. R., Clayton, M. K., Wolff, J. A., & Ney, D. M. (2009). Improved nutritional management of phenylketonuria by using a diet containing glycomacropeptide compared with amino acids. *The American Journal of Clinical Nutrition*, 89(4), 1068–1077.

van Calcar, S. C., & Ney, D. M. (2012). Food Products Made with Glycomacropeptide, a Low-Phenylalanine Whey Protein, Provide a New Alternative to Amino Acid–Based Medical Foods for Nutrition Management of Phenylketonuria. *Journal of the Academy of Nutrition and Dietetics*, 112(8), 1201–1210.

van Spronsen, F. J., & Enns, G. M. (2010). Future treatment strategies in phenylketonuria. *Molecular Genetics and Metabolism*, 99, S90–S95.

van Spronsen, F. J., van Rijn, M., Dorgelo, B., Hoeksma, M., Bosch, A. M., Mulder, M. F., … Verkerk, P. H. (2009). Phenylalanine tolerance can already reliably be assessed at the age of 2 years in patients with PKU. *Journal of Inherited Metabolic Disease*, 32(1), 27–31.

van Spronsen, F. J., van Wegberg, A. M., Ahring, K., Bélanger-Quintana, A., Blau, N., Bosch, A. M., … Giżewska, M. (2017). Key European guidelines for the diagnosis and management of patients with phenylketonuria. *The Lancet Diabetes & Endocrinology*.

van Wegberg, A. M. J., MacDonald, A., Ahring, K., Bélanger-Quintana, A., Blau, N., Bosch, A. M., … van Spronsen, F. J. (2017). The complete European guidelines on phenylketonuria: diagnosis and treatment. *Orphanet Journal of Rare Diseases*, *12*(1).

Vitaflo UK (2017). Vitaflo UK: PKU Sphere GMP product information page and data card. Retrieved from: https://www.nestlehealthscience.co.uk/vitaflo/conditions/protein%20metabolism/phenylketonuria/pku-sphere

Waisbren, S. E., Noel, K., Fahrbach, K., Cella, C., Frame, D., Dorenbaum, A., & Levy, H. (2007). Phenylalanine blood levels and clinical outcomes in phenylketonuria: A systematic literature review and meta-analysis. *Molecular Genetics and Metabolism*, *92*(1–2), 63–70.

Walter, J. H., White, F. J., Hall, S. K., MacDonald, A., Rylance, G., Boneh, A., … Vail, A. (2002). How practical are recommendations for dietary control in phenylketonuria? *The Lancet*, 360(9326), 55–57.

Williams, N. T. (2008). Medication administration through enteral feeding tubes. *American Journal of Health-System Pharmacy*, 65(24), 2347–2357.

Williams, R. A., Mamotte, C. D., & Burnett, J. R. (2008). Phenylketonuria: an inborn error of phenylalanine metabolism. *The Clinical Biochemist Reviews*, *29*(1), 31.

World Health Organisation (2007). Protein and amino acid requirements in human nutrition: report of a joint WHO/FAO/UNU Expert Consultation ; [Geneva, 9 - 16 April 2002]. Geneva: WHO.

Zaki, O. K., El-Wakeel, L., Ebeid, Y., Ez Elarab, H. S., Moustafa, A., Abdulazim, N., … Elghawaby, A. (2016). The Use of Glycomacropeptide in Dietary Management of Phenylketonuria. *Journal of Nutrition and Metabolism*, 2016, 1–5.

The effect of glycomacropeptide-based foods upon blood phenylalanine control in adults and children with phenylketonuria: a systematic review.

0.0: Journal selection

The European Journal of Clinical Nutrition (https://www.nature.com/ejcn/) was selected for publication. Dietary therapies meet its scope, it publishes PRISMA-compliant systematic reviews without requiring prospective protocol registration (something overlooked), decisions regarding acceptance are quick (~47 days) and publication online is rapid (~25 days). The journal is also widely indexed to increase visibility. The journal's low impact factor (2.954) when compared to leading nutrition journals such as the American Journal of Clinical Nutrition (Clarivate, 2018) will reduce visibility but an inconclusive review concerning a rare disease by a student would likely be rejected by a journal like this with a low acceptance rate (20%) and highly competitive publication process (AJCN, 2018). The journal's formatting requirements resemble the university's (e.g. 5000 word limit, 8 table maximum) so only minor adaptations will be required (e.g. referencing).

0.1: Abstract.

Background: Conventional phenylketonuria (PKU) treatment *controls* plasma phenylalanine concentrations (henceforth: plasma phe) by restricting intake of phenylalanine (phe) from protein-containing foods and prescribing amino acid-based 'formula' to meet protein requirements. Glycomacropeptide-based foods form novel alternatives to formula that may improve satiety, acceptability and compliance. However, glycomacropeptide (GMP) contains some phenylalanine raising concerns regarding its impact upon plasma phe in adults and particularly children, who tolerate less phenylalanine.

Objectives: To clarify GMPs effect upon compliance and control in adults and children with phenylketonuria.

Methods: Identical PRISMA-compliant systematic reviews were conducted for adults (over 16) and children (3 – 16). They largely considered any controlled, human study comparing GMP to formula acceptable, including non-randomised studies. Multiple databases, registries, grey literature and snowballing methods were utilised. The Cochrane risk of bias tool and GRADE were adapted to evaluate studies and structured, narrative analyses used.

Results (adults): Six studies were included but conclusions were based upon the sole high-quality randomised trial to avoid bias. It suggests control will deteriorate by ~10% on average if GMP totally replaces formula, but notes marked inter-individual variation in responses (from ~200 µmol/L improvement to ~400 µmol/L deterioration) that remains unexplained. As patients finding GMP acceptable were selectively recruited and allowed to self-select GMP foods, less optimistic responses may occur under less ideal conditions. No compelling evidence of compliance improvements was located and the control decreases described partly related to compliance difficulties as every patient failed to reduce dietary protein intake to accommodate the phe in GMP.

Results (children): The two studies included together suggest formula cannot be totally replaced by GMP in all children without control leaving safe ranges. The amount 'safely replaceable' varied from ~30 – 100%, suggesting control monitoring and adjustments are necessary in practice. However, both studies had extremely limited, heterogeneous designs susceptible to bias, confounding, measurement error and neither monitored compliance adequately. Moreover, both used preliminary GMP formulations with different LNAA (Large Neutral Amino Acid) and phe contents from commercial products, something likely to decrease effectiveness.

Conclusions: The review was unable to describe GMPs effect upon compliance and control with any certainty. Caution is thus advised among practioners choosing to prescribe GMP, particularly to children, until additional randomised trials are conducted to clarify matters. If GMP is utilised, products high in LNAA and low in phe should be selected (e.g. Glytactin), control and compliance should be monitored carefully and in children practioners should be prepared to adjust the amount of GMP introduced if control leaves safe ranges.

1.0: Introduction.

Phenylketonuria impairs Phenylalanine Hydroxylase functionality, causing Phe to accumulate to toxic levels in plasma and tissues (Blau, van Spronsen & Levy, 2010). Prior to the development of dietary treatment, PKU was associated with severe neurocognitive disabilities usually requiring residential care (Paine, 1957). Dietary treatment is initiated at birth and restricts dietary phenylalanine to *control* plasma phe (van Wegburg et al., 2017). Overwhelming evidence supports its effectiveness (Waisbren et al., 2007; Fonnesbeck, McPheeters, Krishnaswami, Lindegren & Reimschisel, 2012; Albrecht, Garbade & Burgard, 2009) and since its widespread adoption the disabilities previously commonplace have become rare and most treated patients lead normal lives (Koch et al, 2002).

However, treatment has limitations. As phenylalanine is ubiquitous in proteins most protein-containing foods must be restricted, even staples (van Wegburg et al., 2017). Moreover the phe-free protein substitutes (formula) prescribed to address protein requirements lack acceptability & convenience (MacDonald, Harris, Rylance, Asplin & Booth, 1997; Macdonald, 2000), limitations that improvements have not totally resolved (Prince, McMurry & Buist, 1997; Macdonald et al., 2004). As a result, adherence proves difficult in all ages (MacDonald, Gokmen-Ozel, van Rijn & Burgard, 2010; Walter et al., 2002) particularly

with formula (Schulz & Bremer, 1995; Prince et al., 1997). As adherence improves control (Macdonald et al., 2006; Mönch, Herrmann, Brösicke, Schöffer & Keller, 1996) and control predicts neurocognitive outcomes (Waisbren et al., 2007; Fonnesbeck et al., 2012) these limitations of treatment likely contribute to poor control and below-normal neurocognitive outcomes remaining commonplace among early-treated patients (DeRoche & Welsh, 2008; Enns et al., 2010). They also explain interest in alternative treatments (van Spronsen & Enns, 2010; Specola & Chiesa, 2017) including Glycomacropeptide.

Glycomacropeptide (GMP), the only naturally phe-free intact protein, is used to manufacture foods forming alternatives to formula (Van Calcar & Ney, 2012). Emerging evidence suggest these have the potential to improve the palatability (Ney et al., 2016; Lim, van Calcar, Nelson, Gleason & Ney, 2007; Van Calcar et al., 2009; Zaki et al., 2016) and satiety (MacLeod, Clayton, van Calcar& Ney, 2010) of PKU diets and may also improve bone health (Solverson, Murali, Litscher, Blank & Ney, 2012) and gastrointestinal comfort over formula (Ney et al., 2016). As patients attribute compliance difficulties to the poor palatability and satiety of formula, compliance and control improvements may be expected to ensue. However, as purification methods are imperfect, GMP contains residual phenylalanine (La Clair, Ney, Macleod & Etzel 2009) raising concerns regarding its impact on control, particularly among children that are more susceptible to phe. Despite GMP products having UK approval for children (Cambrooke, 2018; Vitaflo, 2017) the most recent evidence-based guidelines acknowledge that these concerns remain unresolved (van Wegberg, 2017).

A non-systematic review was conducted to clarify matters but was unsuccessful. It located three studies comparing GMP to formula in adults (over 16). One reported improved control

(by 10%) and compliance on GMP (Ney et al., 2009). Another (van Calcar et al., 2009) reported equivalent control, a supportive finding given GMP's other advantages. However, both had limited designs. The former was a case study meaning the subject may be an outlier. The latter was unrealistically short, lacked randomisation and used an inpatient setting, preventing compliance differences from occurring. The strongest study located was randomised and (conversely) reported ~10% worse control on average alongside compliance difficulties (Ney et al., 2016). This suggests bias may explain previous supportive findings and raises concerns regarding GMP's impact upon compliance and control in adults. As findings were conflicting, these concerns require clarification, particularly as the review did not use pre-determined methods to review the entire evidence base in an unbiased manner or formally evaluate the contribution of heterogeneity or bias to results (Mulrow, 1994; Oxman & Guyatt, 1993) as a PRISMA-compliant systematic review would (Liberati et al., 2009). Existing GMP reviews have similar limitations (Van Calcar & Ney, 2012; Ney, Blank & Hansen, 2014; Ney & Etzel, 2017), come from one group with commercial interests in GMP raising objectivity concerns (Oxman & Guyatt, 1993) do not specifically consider compliance or control and require updating. A PRISMA-compliant systematic review was thus performed to clarify GMPs effect on control (objective 1a) and compliance (objective 1b) in adults.

The review located three studies involving children (3 – 16). Whilst control findings were conflicting, heterogeneous approaches to replacing formula with GMP were used, making comparisons difficult. Abdel-Salam & Effat, (2010) reported questionable control improvements (30 – 80% over 3 days) and poor reporting left the amount of formula replaced unclear. Zaki et al., (2016) replaced 50% of formula and reported equivalent

control. Daly, Evans, Chahal, Santra & MacDonald, (2017) adjusted the amount of formula replaced on an individual basis to prevent control leaving safe ranges for ethical reasons, and replaced 30 – 80%. Furthermore, all had non-randomised, bias-prone designs, unclear reporting and pilot-tested incomplete GMP formulations poorly representing those used clinically. As only three studies with conflicting findings and heterogeneous, low quality designs were located, GMPs effect upon compliance and control in children is poorly understood, despite control being crucial (van Wegberg et al., 2017). An identical systematic review was thus performed to clarify the effect of GMP upon control (objective 2a) and compliance (objective 2b) in children.

2.0: Methods.

2.1: Overall approach

Methods were selected using guidance from the Cochrane Collaboration (Higgins, Churchill, Chandler & Cumpston, 2017) and Centre for Reviews & Dissemination (CRD, 2009). Reporting met PRISMA requirements (Moher et al., 2009) and a PRISMA-P compliant protocol was prepared prospectively (appendix 3) to ensure transparency (Shamseer et al., 2015).

2.2: Eligibility criteria

Given the few, low-quality studies previously located, eligibility criteria (tables 19-20) favoured inclusivity. Most controlled human studies comparing GMP to formula were eligible, including non-randomised studies (NRS), despite the likelihood of this increasing heterogeneity and bias.

2.3: Search Strategy

For details of resources searched & terms used see tables 12-18 (appendix 1). A librarian assisted with search strategy development. To increase sensitivity, multiple databases were utilised alongside snowballing methods including hand searches of key journals (Suarez-Almazor, Belseck, Homik, Dorgan & Ramos-Remus, 2000), reference list searches of key articles and several citation search tools. To counteract publication bias, unpublished literature and registries were included (Song, Eastwood, Gilbody, Duley & Sutton, 2000) but as translation facilities were unavailable non-English studies were excluded, risking language bias (Egger et al., 1997). A PubMed search strategy (table 12) was devised first by adapting Cochrane's IEM strategy (Cochrane, 2018). To favour sensitivity, only the condition (PKU) and intervention (GMP) aspects of PICOS were combined using 'AND', but many synonyms for each were combined using 'OR' (CRD, 2009). Natural language was preferred over database-specific terms to expediate adaptation for other databases and because these can be mis-applied (Relevo & Balshem, 2011). The strategy was pilot-tested, adapted for other resources and checked against peer-reviewed guidance (McGowan et al., 2015).

2.4: Data management, screening & data extraction

Data was managed using EndNote, tables 12-18 and a PRISMA flowchart (figure 1). After duplicate removal, abstracts and titles of studies located using database-type resources (tables 12-14) were screened (table 19). Screened studies then directed searches of snowball-type resources (tables 15-18). Different reports of the same study were next merged but retained to check for selective reporting (CRD, 2009). Remaining full text articles

were checked against inclusion criteria (table 20) and exclusions explained (table 7). Data required to address objectives, describe methods and evaluate quality was extracted using standard forms (table 21). Authors were emailed regarding missing data (table 22) and allowed 18 days to respond though studies were included irrespective of responses. All forms were developed using PICOS (Participants, Intervention, Control, Outcome, Study design) a-priori to minimise bias, piloted and their completion was double-checked to reduce errors as repetition by another reviewer (Higgins & Deeks, 2011) was unavailable.

2.5: Quality evaluation of individual studies

The Cochrane risk of bias tool (Higgins, Altman & Sterne, 2017) was adapted to evaluate study quality. Unlike numerical scales and checklists (Crowe & Sheppard, 2011; Moher et al., 1995) this ensures reviewers transparently explain decisions, consider the importance of each bias form to their review's topic (table 2), consider the likely magnitude and direction of bias and distinguish conduct (which risks bias) from reporting (Higgins et al., 2011). The tool was adapted to incorporate within-subjects designs (common in PKU) via a carryover domain (Higgins, Deeks & Altman, 2011). Non-randomised studies were considered at risk of selection bias irrespective of the apparent distribution of confounding variables or attempts to balance them (Deeks et al., 2003) but otherwise evaluated like randomised studies (RS) (Reeves, Deeks, Higgins & Wells, 2011). The tool was expanded to evaluate aspects of design quality not relating to bias such as sample size & precision (which quantitative meta-analyses incorporate), measurement errors and external validity as these also reduce confidence in findings (CRD, 2009).

2.6: Data analysis

Given the bias-prone, heterogeneous studies previously located, quantitatively pooling results may have attached unwarranted credibility to findings and mislead readers (CRD, 2009; Deeks, Higgins & Altman, 2017). Narrative analyses were instead performed using structured approaches (CRD 2009, page 45; Popay et al., 2006). Separate analyses were performed for adults (over 16) and children (3-16) since phe tolerance, protein and formula requirements differ (van Wegberg et al., 2017) and totally replacing formula with GMP appears impossible in some children (Daly et al., 2017). Quality assessments were incorporated into analyses through tables (3 & 8) comparing effect estimates to study limitations. To avoid misleading findings due to bias randomised/ high quality studies were analysed separately and conclusions based primarily on them, where possible (Reeves et al., 2011). The contribution of heterogeneity to results was evaluated using narrative subgroup-analyses which grouped studies by potentially mediating characteristics and compared effects (Popay et al., 2006). Possible mediators of responses were listed a-priori though no hypotheses were formed regarding them.

2.7: Review-wide quality evaluation.

The confidence that can place in conclusions was rated using GRADE (Guyatt et al., 2011). Risk of publication bias was evaluated subjectively as compatible precision data for funnel plots/ statistical tests was unavailable due to poor reporting. The approach recreated a funnel plot by comparing the number of small, industry-linked, supportive studies and small unsupportive studies (CRD, 2009).

3.0: Results

Figure 1: PRISMA flowchart of review process.

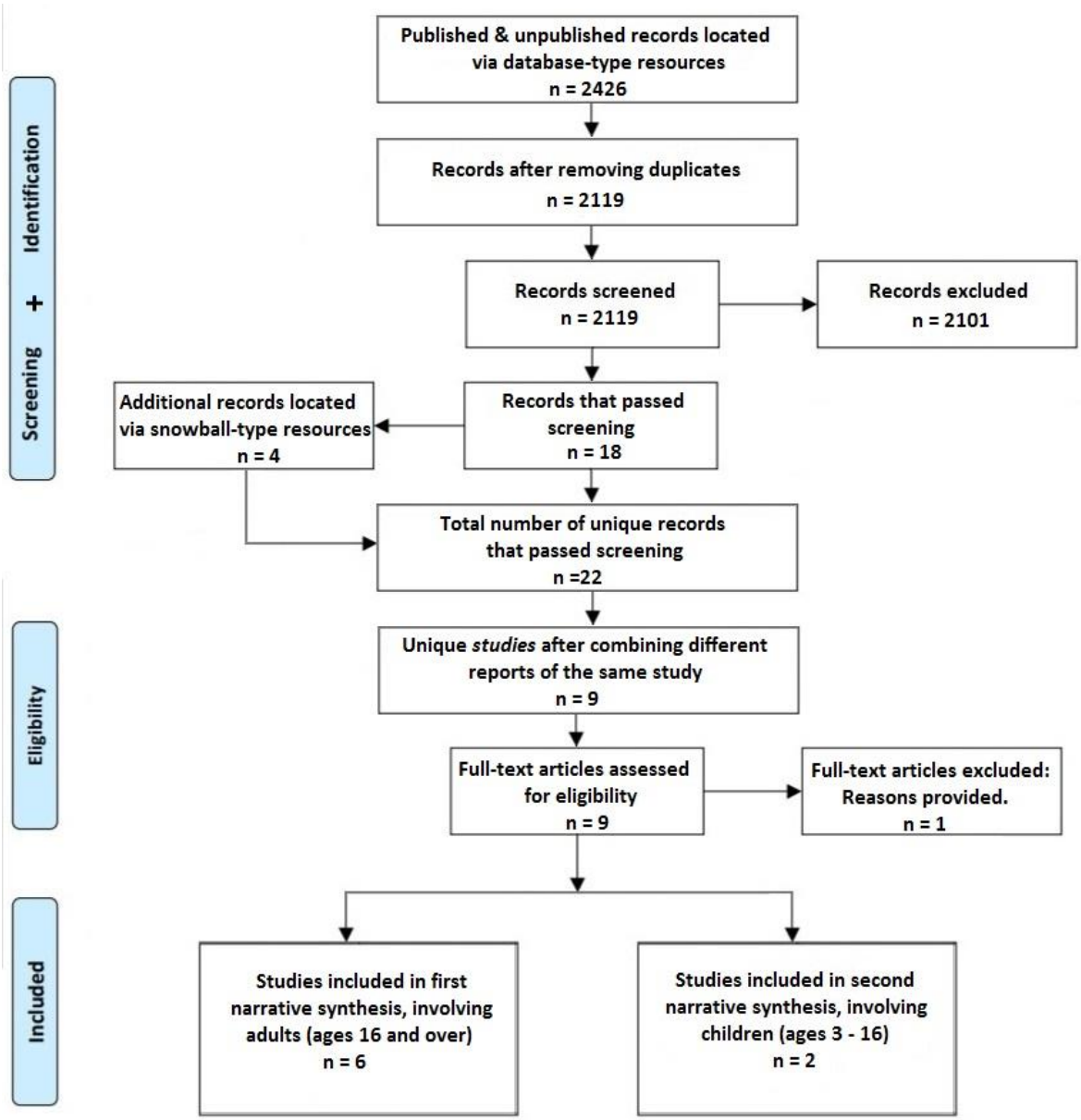


Table 1a: Characteristics of included adult studies.

Study	Methods	Participants	Intervention (GMP) & control (AA)	Measurements taken	Notes
LaClair et al., (2009)	Early study primarily investigating the purification and supplementation of GMP to form a complete Amino acid (AA) source. Secondarily monitored control in a subset of n = 4 (from 15) subjects sequentially undergoing 4 days usual treatment (formula) then 4 days GMP treatment in a non-randomised, self-controlled, within-subjects design.	**Number:** 4 **Ages:** 19-29 **Population:** unknown **Severity:** unknown **History:** unknown **Setting:** outpatient **Country:** USA	**GMP formats:** Pudding (food-like) **Phe content:** unknown **LNAA content:** High vs formula (150% of USA Recommended Daily intake (RDI) for Tyrosine, 130% for other limiting LNAA) **Nutritionally complete:** Yes **Commercial product:** No **% formula replaced:** 100% **Exchanges removed?** Yes, diets contained equal phe **Same GMP in every patient:** yes **AA formats:** Unknown **Same formula in every patient:** No	**Timepoints measured:** Single, post prandial measurement 2.5h after breakfast on final 2 days of each 4 day condition. **Collection and analysis:** Plasma samples obtained via venepuncture, analysed via Beckman 6300 amino acid analyser (ion exchange chromatography).	Many aspects of design remain unclear as published report lacked detail and no response was received to request for clarification.
Ney et al., (2009)	Non-randomised experimental case study comparing usual treatment (formula) to GMP over 15 weeks in n = 1 outpatient. Formula was consumed for the first 3 weeks and last 2 weeks, and GMP for the middle 10 weeks but data is reported for 6 weeks during which weighed, portioned food of known phe content was provided (2 weeks formula & 4 weeks GMP).	**Number:** 1 **Age:** 29 **Population:** clinic patient **Severity:** Classical **History:** on-diet but lapsed as adolescent **Setting:** outpatient **Country:** USA	**GMP formats:** sports drink, milkshake, puddings, snack bar (food-like) **Phe content:** 0.4g/ 100g GMP **LNAA content:** As above **Nutritionally complete:** Yes **Commercial product:** No **% formula replaced:** 100% **Exchanges removed?** Yes, diets contained equal phe **Same GMP in every patient:** yes, n=1 **AA format:** weighed powder **Same formula in every patient:** n=1	**(a) Plasma samples:** collected via venepuncture (fasting) between 12:00 to 12:30 at local clinic. 4 samples collected over GMP and formula conditions. Analysed via Beckman 6300 AA analyser. **(b) Bloodspots:** collected by patient from 9.00-9.30 (fasting). 8 samples collected in GMP condition, 4 in AA condition. Analysed via tandem Mass spectrometry.	Full details available as response was received to request for clarification.

Table 1b: Characteristics of included adult studies (continued).

Study	Methods	Participants	Intervention (GMP) & control (AA)	Measurements taken	Notes
van Calcar et al., (2009)	Inpatient metabolic study with non-randomised, within-subjects design. n=11 patients underwent 4 days usual treatment (AA) then 4 days GMP treatment in a non-randomised order. Subjects were inpatients for days 2-4 of AA treatment and all four days of GMP treatment.	**Number:** 11 **Age:** 11-31 (n = 3 < 16) **Population:** clinic patients **Severity:** 10 Classical, 1 variant (>1000 µmol/L) **History:** early treated. No other details. **Setting:** outpatient **Country:** USA	**GMP formats:** sports drink, milkshake, pudding, snack bar (food-like) **LNAA content:** High vs formula (150% of USA RDI for Tyr, 130% for other limiting LNAA) **Phe content:** As above **Nutritionally complete:** Yes **Commercial product:** No **% formula replaced:** 100% **Exchanges removed?** Yes, diets contained equal phe **Same GMP in every patient:** no **AA format:** mainly weighed powder **Same formula in every patient:** no	Data presented compares postprandial (n=11) and fasting (n=6) plasma phe data for final day of each condition (days 4 vs 8). Analysis is via Beckman 6300 AA analyser. Fasting data collected only in subset of n=6 patients. Data was collected for the final 2 days of each condition (3 vs 7 & 4 vs 8) but data from days 3 & 7 were omitted due to space constraints in the journal. This does not affect findings.	Full details available as response was received to request for clarification.
Ney et al., (2016)	Randomised, two-centre crossover trial comparing control in n=30 free living outpatients during 3 weeks usual treatment (formula) and 3 weeks GMP treatment, separated by a 3 week washout on formula.	**Number:** 30 **Age:** 15 - 49 **Population:** PKU patients at 2 centres **Severity:** 20 Classical, 10 variant. **History:** early treated. optimal control not required. **Setting:** Outpatient **Country:** USA	**GMP formats:** Food-like (e.g. pudding). **Phe content:** 1.8 mg/ g PE (foods) **LNAA content:** As above **Nutritionally complete:** Yes **Commercial product:** Yes, Glytactin. **% formula replaced:** 100% **Exchanges removed?** Yes, diets designed to contain equal phe **Same GMP in every patient:** no, self-selected favourite format **AA format:** mix of food-like and weighed powder. Some took both. **Same formula in every patient:** no.	(a) For the main effect, 4 fasting venous blood samples were taken during study visits: (1) baseline pre-AA (2) end of AA (3) baseline pre-GMP (4) end of GMP. Analysed via Hitachi L-8900 amino acid analyser. (b) Secondary analyses used dried blood spots taken at 8 points during each 3 week treatment period (on days 3, 6, 7, 10, 13, 14, 17, 21). Analysed via tandem MS.	Full details available as response was received to request for clarification. **NB.** Five subjects (with variant PKU) completed the study whilst taking Sapropterin treatment. However, their phe tolerance was stable at baseline and the Sapropterin dose remained consistent throughout.

Table 1c: Characteristics of included adult studies (continued).

Pinto et al., (2017)	Observational study that retrospectively compares longitudinal control data from n=11 patients undergoing periods of AA (13 ± 5m) and GMP treatment (13 ± 7 m) during routine clinical practice. The amount of formula replaced by GMP varied (27-100%) according to choice or assignment by clinicians: mean 57% (27 – 100%).	**Number:** 11 **Age:** Mean = 27 (n=1 <16: 13 years old). **Population:** Struggling to maintain control. **Severity:** 1 HPA, 4 mild, 6 classical **History:** Unknown **Setting:** Outpatient **Country:** Portugal	**GMP formats:** Mainly weighed powder **Phe content:** 1.7mg/ g PE (foods) **LNAA content:** As above **Nutritionally complete:** Yes **Commercial product:** Yes, Glytactin. **% formula replaced:** 27-100% **Exchanges removed?** No. GMP diet contained extra phe (34 ± 12 mg) **Same GMP in all patients:** all but one **AA format:** mix of food-like and weighed powder. Some took both. **Same formula in every patient:** no.	Control monitored via fasting blood spot collection: **i) Formula comparator:** median of 11 blood spots taken from May 2013 until GMP introduction (13 ± 5 months) **ii) GMP intervention:** median of 15 blood spots taken from GMP introduction until the second evaluation (13 ± 7 months). Analysed via tandem MS.	
Ahring et al., (2018)	Short-term, single blinded, randomised crossover trial comparing the effect of four drink mixtures (consumed with a standard meal) on postprandial plasma phe for 240 minutes after ingestion in n=8 subjects. **Comparison 1:** Pure GMP (Lacprodan CGMP-20) vs formula with the same AA profile. **Comparison 2 (of interest):** GMP supplemented with limiting AA to form complete protein source vs formula with the same AA profile, minus phe.	**Number:** 8 **Age:** 16 – 48 (mean 33) **Population:** Recruited via national database **Severity:** all Classical **History:** early & continuously treated. **Setting:** Attended centre 4 times (for each drink) **Country:** Denmark	**GMP format:** drink compositions in first column **Phe content:** GMP = 0.16 g Phe/ 100 g AA **LNAA content:** Unclear **Nutritionally complete:** Drink 3 was. **Commercial product:** No. **% formula replaced:** 100% for 1 meal **Exchanges removed?** No, only investigated 1 meal **Same GMP in every patient:** yes (for 1 meal) **AA format:** drinks, see first column **Same formula in every patient:** yes	On each of four visits (one for each drink) venous blood was collected as follows: (a) fasting samples taken in the morning (0 minutes, baseline) (b) the meal was consumed (c) postprandial samples were taken 15, 30, 60, 120, and 240 minutes after the meal. Analysis was via HPLC-MS/ MS	Some details unknown as no response was received to a request for clarification.

y = year, m = month, HPLC = high performance liquid chromatography, MS = mass spectrometry

Table 2a: Identification of important quality domains.

Domain	Importance	Explanation	Likely direction and magnitude of bias
Random sequence generation (1)	High	In parallel designs, prevents selection bias I.E. systematic differences in the distribution of prognostic factors between groups which may confound changes in control otherwise attributable to the intervention or comparator (Higgins et al., 2017). Within-subjects designs are common in PKU to control for between subject differences (e.g. in severity). In these, randomisation instead prevents confounding from order effects and trends in outcomes over time. It is unclear how these would occur in PKU but only adequate randomisation accounts for unknown factors (Reeves, Deeks, Higgins & Wells 2011).	**Generally overstates benefit of GMP.** Varies (Odgaard-Jensen et al., 2011) but results are generally overly-optimistic where randomisation is inadequate as subjects likely to respond well tend to be preferentially assigned/ select the intervention or a treatment order (Savovic et al., 2012).
Concealment of allocations (2)	High	Serves the same purpose as random sequence generation. Personnel aware of upcoming allocations may preferentially assign subjects to particular treatments (parallel studies) or treatment orders (within-subject studies). Concealment of allocations prevents this. It is distinct from blinding which occurs after randomisation..	**Generally overstates benefit of GMP.** As above.
Blinding of participants & Personnel (3)	**Low**	Prevents participants and personnel gaining awareness of allocations to prevent performance bias. I.E. systematic differences between conditions in the care provided or exposure to confounding factors. Blinding of participants is unimportant as GMP tastes different from formula and the possibility of this promoting compliance improvements is being investigated. Blinding of personnel is also unimportant. It is difficult as GMP and formula look different and unlikely to directly affect control, an objective outcome. It may affect control indirectly by promoting compliance but this would occur in practice.	**Generally overstates benefit of GMP.** Meta-epidemiology suggests various forms of unbinding are generally associated with overly optimistic effect estimates, particularly where outcomes are subjective (Savovic et al., 2010). Analyses for separate sources of unbinding are not available.
Blinding of outcome assessment (4)	**Low**	Personnel aware of allocations may measure outcomes in a manner that introduces bias. This is not important as objective measures like plasma phe are far less subject to manipulation, misinterpretation and bias than subjective measures such as pain.	**Generally overstates benefit of GMP.** As above. Bias is less likely and of smaller magnitude with objective measures like phe.
Incomplete outcome data (5)	High	Systematic differences between conditions in the number of exclusions or withdrawals can bias results if the reasons underpinning them relate to prognostic factors/ outcomes and they are handled improperly. E.g. patients disproportionately withdrawing from the GMP condition as they perceive no benefit or are struggling to comply may overstate GMPs benefit if their data are excluded, as good responders will be over-represented in analyses (Higgins et al., 2017; CRD, 2008).	**Varies.** Depends entirely on the extent of withdrawals/ exclusions, when they occurred, their distribution between conditions, the underlying reasons and how incomplete data is handled.
Selective outcome reporting (6)	High	Selectively publishing supportive findings e.g. those attaining statistical significance or supportive sub-sets of data is common and can bias results as non-supportive findings are equally important (Dwan, Gamble, Williamson & Kirkham, 2013). The review contains many non-randomised studies which are inherently at risk of selective outcome reporting as they cannot be checked against protocols, which are not required (Higgins et al.,, 2017).	**Generally overstates benefit of GMP.** (Dwan, Gamble, Williamson & Kirkham, 2013).

Table 2b: Identification of important quality domains (continued).

Domain	Importance	Explanation	Likely direction and magnitude of bias
Carryover in within-subject designs (7)	Low	Carryover is a form of bias unique to serial, within-subjects designs caused by the effects of earlier treatments persisting into periods when subsequent treatments are administered and causing confounding (Higgins et al., 2011). Statistical tests for carryover are underpowered so this review assumes carryover is not present unless studies provide reasons to expect it (Mills et al., 2000). Barring these, carryover is unlikely as PKU is a stable, chronic condition that cannot spontaneously improve or deteriorate and protein substitutes have short-lived effects upon plasma phe that rapidly reverse upon cessation.	**Generally overstates benefit of GMP.** Protein substitutes lower plasma phe so inadvertently combining the effects of two would likely improve control.
Influence of funding source (8)	High	Funding providers or personnel with vested interests in the outcomes of studies may introduce bias, for example by permitting only selected results to be published. However, industry funding does not necessarily introduce bias so specific evidence of bias is needed.	**Generally overstates benefit of GMP.**
Validity & reliability of Outcome Assessment (9)	High	If invalid and unreliable methods are used to measure or compare outcomes, changes attributed to the intervention or comparator may be partly attributable to measurement error. In this instance the difference between most modern methods are inconsequential but dried blood spots analysed by tandem MS recover 8-28% less phe than criterion methods and semi-quantitative methods like the original Guthrie Assay are sometimes used (van Wegberg et al., 2017).	**Varies.** Depends entirely on the extent and distribution of errors.
External validity of study (10)	High	Various factors can reduce the generalisability of findings to clinical practice (a) Some studies use weak comparators by comparing powdered formula requiring weighing & preparation to food-like GMP products, which may exaggerate compliance/ control differences as ready-to-drink formulas are increasingly used (b) Inpatient studies do not recreate practice and all report 100% compliance, suggesting Hawthorne effects mask important compliance differences (c) some studies selectively recruit patients that like GMP & allow them to self-select from a wide range of foods, something unlikely to occur in practice. (d) some studies use preliminary forms of GMP poorly representing commercial products e.g. by lacking LNAA or containing too much phe.	**n/a. Influences external validity.** Strictly speaking, external validity depends on the scenario findings are applied to. However weak comparators, inpatient studies and selective recruitment are likely to overstate benefit. Using preliminary forms of GMP is likely to understate benefit as LNNA promote control and additional phe impairs it.
Sample size and statistical analyses (11)	High	Quantitative meta-analyses attribute more importance to studies with larger samples as these increase power & precision by better accounting for variation between subjects and reducing the possibility of chance explaining findings. Narrative analyses do not unless explicit attempts are made to do so. The appropriateness of statistical analyses also influence how much weight should be attached to findings.	**n/a. Influences precision and power.** Small samples and wide confidence intervals reduce the confidence that can be placed in findings as chance or random error may explain them.
Other concerns RE bias & quality (12)	High	A catchall for threats to validity from design limitations, sources of bias or confounding not noted elsewhere. Examples include claims that the study is fraudulent and study-specific design limitations like recall bias in retrospective designs.	**Varies.** Depends on form of bias or limitations identified.

Table 3: Comparison of effect estimates and design quality: adult studies.

Study	Number of patients	**Effect estimate.** (mean difference in plasma phe between GMP and formula conditions. Values are mean ± SE in µmol/L in fasting condition unless stated).	**Did control improve? (compared to formula)**	1. Random sequence generation (selection bias/ order effects)	2. Concealment of allocations (selection bias/ order effects)	3. Blinding of participants & Personnel (performance bias)	4. Blinding of outcome Assessments (detection bias)	5. Incomplete outcome data (attrition bias)	6. Selective outcome Reporting (reporting bias)	7. Carryover in within-subjects Designs (carryover bias)	8. Influence of funding source (may introduce several biases)	9. Validity & reliability of outcome assessment (measurement error)	10. Reduced external validity	11. Sample size and statistical analyses (low power & precision)	12. Other threats to validity	% of domains at low risk	% of important domains at low risk
Ney et al., (2016)	30	GMP: 62 ± 40 increase (non-sig) AA: 85 ± 40 decrease (sig) **GMP: 147± 39 higher (p< 0.05)**	✘ Slightly Worse (10%)	Low Risk	Low Risk	High Risk	High Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	High Risk	Low Risk	Low Risk	**75**	**89**
----------Higher quality study above. Lower quality studies below----------																	
van Calcar (2009)	11	GMP: 676 ± 92 AA: 619 ± 82 (postprandial) **No sig difference (p >0.05)**	✔ Equivalent	High Risk	High Risk	High Risk	High Risk	Low Risk	Uncl Risk	Low Risk	Low Risk	Low Risk	High Risk	Low Risk	Low Risk	**50**	**56**
Ahring et al., (2018)	8 (6)*	Comparing drinks 3 & 4* at 15 to 240 mins (postprandial) **No sig differences (p >0.05)**	✔ Equivalent	Low Risk	Uncl Risk	High Risk	High Risk	Low Risk	Uncl Risk	Low Risk	Low Risk	Low Risk	High Risk	High Risk	Low Risk	**50**	**56**
Ney et al., (2009)	1	GMP: 667 ± 24 AA: 736 ± 13 **GMP: 10% lower (p <0.05)**	✔✔ Slightly Better (10%)	High Risk	High Risk	High Risk	High Risk	Low Risk	Uncl Risk	Low Risk	Low Risk	Low Risk	High Risk	High Risk	Low Risk	**42**	**44**
La Clair et al., (2009)	4	GMP: 400 ± 200 AA: 400 ± 300 (mean ± SD) **No sig difference (p >0.05)**	✔ Equivalent	High Risk	High Risk	Uncl Risk	Uncl Risk	Uncl Risk	Uncl Risk	Low Risk	Low Risk	Low Risk	High Risk	High Risk	Low Risk	**33**	**33**
Pinto et al., (2017)	11	GMP: 521 ± 139 AA: 545 ± 133 (mean ± SD) **No sig difference (p >0.05)**	✔ Equivalent	High Risk	High Risk	High Risk	High Risk	Low Risk	Uncl Risk	Low Risk	Low Risk	High Risk	Low Risk	High Risk	High Risk	**33**	**33**

Abbreviations: Uncl = unclear, SE = standard error. *only 6 subjects participated in the comparison of interest: Complete GMP vs AA with identical composition, minus phe

Table 4: Compliance during GMP treatment in adults.

Study	**Compliance with protein substitute prescriptions (overall and frequency)**	**Compliance with other dietary requirements.**	**Where valid & reliable instruments used?**	**Was compliance clearly improved in the GMP condition?**
Ney et al., (2016)	**Overall consumption:** (a) No significant difference reported in medical food logs: Lower than prescriptions by ~10 g PE / kg BM /d in both conditions (b) more objective plasma threonine measurements detected 6 patients complying poorly with GMP (no equivalent measure for formula) **Frequency of consumption:** No different overall. Improvement noted for GMP in a sub-analysis comparing only patients receiving GMP & formula second.	**Dietary phe intake increased significantly (p = 0.0259) in the GMP condition** (by 88 ± 6 mg Phe/d) despite the diets being designed to contain equal phe content, as subjects did not remove dietary phe to accommodate the additional phe introduced by GMP as required.	**Overall:** Yes. **Protein subs:** Daily logs & plasma threonine (objective measure) **Diet:** Food diaries for the last 3 days of each 3 week each condition. Some estimated, some weighed. Eact numbers estimated and weighed unclear.	**No.** Compliance with protein substitutes was either equivalent or worse (despite a sub-analysis reporting improved frequency for GMP when comparing only patients receiving GMP and formula second, which has no logical basis). Dietary compliance was worse in the GMP condition as patients did not compensate for the additional phe introduced by GMP as required.
Ney et al., (2009)	**Overall consumption:** No significant difference between conditions. **Frequency of consumption:** Greater for GMP (3 times per day) than AA (once per day). Attributed to better taste.	**No differences noted between conditions** (emailed author to confirm).	**Overall:** Yes. **Protein subs:** Daily logs. **Diet:** Estimated food diaries for last 3 days of each condition & weekly phone interviews.	**Yes.** Greater frequency of ingestion was noted for GMP (3 times per day) than formula (once per day). The sole patient consumed all their daily formula prescription at once due to get rid of it due to its poor taste
van Calcar et al., (2009)	100% compliance in both conditions. For days 1-2 of formula condition, weighed food was provided and uneaten food returned. Other days as inpatients.	100% compliance in both conditions, as previous column.	Apparently, though monitoring methods unclear.	**n/a**. The inpatient setting prevented compliance differences from occurring.
Ahring et al., (2018)	100% compliance in both conditions. Single meal provided in laboratory.	100% compliance in both conditions. Single meal provided in laboratory.	See previous column.	**n/a**. The inpatient setting prevented compliance differences occurring.
La Clair et al., (2009)	Unclear. Clarification requested via email but no response received.	Unclear. Clarification requested but no response received.	Unclear. See previous columns.	**n/a.** Compliance levels not reported and clarification not received via email.
Pinto et al., (2017)	Compliance with protein substitutes was not monitored (emailed author).	Compliance with dietary advice was not monitored (emailed author).	Compliance not monitored.	**n/a.** Compliance was not monitored (emailed author to confirm).

Abbreviations: PE = protein equivalent. BM = body mass. D = day. Subs = protein substitutes.

Table 5: Characteristics of included child studies.

Study	Methods	Participants	Intervention & comparator	Outcomes	Notes
Zaki et al., (2016)	Non-randomised feasibility study with a self-controlled, within-subjects design. n= 10 children underwent 9 weeks GMP treatment (50% GMP, 50% formula) then 9 weeks usual treatment (100% formula, 0% GMP). Data is presented for final 7 weeks to avoid carryover. A preliminary form of GMP cheese-spread was developed and pilot-tested in the study.	**Number:** 10 **Age:** 4 – 16 (median 7) **Population:** Clinic patients **Severity:** all Classical **History:** Compliant for 2m before study. **Setting:** Outpatients **Country:** Egypt	**GMP format:** Spread made from commercial GMP, butter & salt **Phe content:** 4.8mg / g PE (high!) **LNAA content:** Unclear **Nutritionally complete:** unclear **Commercial product:** No. **% formula replaced:** 50% **Same GMP in every patient:** yes **AA format:** Unknown **Same AA in all subjects:** unknown	Phe measured weekly for 7 weeks. Analysis via "tandem mass spectrometry and nonderivatised amino acid/ acyl-carnitine kits" Timing of tests, whether they were fasting and whether blood spots or blood draws were used are all unknown.	Many aspects of design remain unclear as published report lacked detail and no response was received to request for clarification.
Daly et al., (2017)	Non-randomised, prospective, pilot study with parallel design. Compares control over 6 months between n=9 children remaining on 100% formula and n=12 (different) children replacing it with GMP. In 7 subjects, the amount of formula replaced by GMP had to be reduced to prevent control loss on ethical grounds. The amount of formula safely replaceable by GMP was 30% - 100% (median 50%). Again, a preliminary form of GMP was pilot-tested.	**No:** 21 (1 exclusion) **Age:** 6–16 (median 11) **Population:** Clinic patients **Severity:** Varied. See table 2 **History:** 70% of past 2y monitoring optimal. **Setting:** Outpatients **Country:** England	**GMP format:** Unknown **Phe content:** 1.5mg / g PE **LNAA content:** Low compared to formula and GMP used elsewhere. AA profile met WHO 2007 PKU guidelines. **Nutritionally complete:** Yes **Commercial product:** No. **% formula replaced:** 30 – 100%. **Same GMP in every patient:** no **AA format:** Mainly liquid pouches **Same AA in all subjects:** no	Dried blood spots collected once per week after overnight fast. Sent to lab and analysed via tandem MS. Above compared to (a) baseline (b) concurrent control group (between-subjects comparison) (c) retrospective clinical data for previous 12M (within-subjects comparison)	No response received to request for clarification. Retrospective, within-subjects comparison used as between-subject comparisons do not allow for differences in tolerance, severity etc. but a concurrent control group was present.

y = year, m = month, WHO = World Health Organisation. NB phe content refers to content of raw GMP (unless otherwise stated). Content of complete products or meals varie

Table 6: Characteristics of ongoing child studies.

Study	Methods	Participants	Intervention & comparator	Outcomes	Reason not included
Daly, Chahal, Evans & MacDonald (2016)	Continuation of the study described in Daly et al., (2017) which was a pilot for an ongoing 3 year study. Originally, control loss necessitated individual-level adjustments of the amount of formula replaced by GMP and as little as 30% was safely replaceable in some cases (but an average of 50%). In response, GMP was modified as below and the study continued using the same design and subjects. (a) extra tyrosine for 4 months. (b) extra LNAA for 5 months. Results discussed in discussion.	**No:** 21 (1 exclusion) **Age:** 6–16 (median 11) **Population:** Clinic patients **Severity:** Varied. See table 2 **History:** 70% of past 2y monitoring optimal. **Setting:** Outpatients **Country:** England	**GMP format:** Unknown **Phe content:** 1.5mg / g PE **LNAA content:** Base GMP in Daly et al., described above. Drinks b in this extension contained extra LNAA. Quantity unclear (not published). **Nutritionally complete:** yes. **Commercial product:** No. **% formula replaced:** varied. Specifics unknown (not published) **Exchanges removed?** Unknown. **Same GMP in every patient:** no **AA format:** mainly liquid pouches **Same AA in all subjects:** no.	Presumably the same as Daly et al., (2017) above, as this is an extension of the same study. However this not known as the extension is not yet published.	Could not be included as full text and detailed findings unavailable so quality evaluation was impossible. Paper outlining 12 month findings currently undergoing publication. Study initially identified via conference abstract (Daly et al., 2016). Additional details obtained via email correspondence with author.

Table 7: Characteristics of excluded child studies.

Study	Methods	Participants	Intervention & comparator	Outcomes	Reasons for exclusion
Abdel-Salam & Effat (2010)	Non-randomised study comparing control after 3 days GMP treatment to baseline. What baseline treatment consisted of is unclear. Control data is presented for three groups but the number of patients in each and their ages are also unclear.	**Number:** 10 or 30: unclear **Age:** "infants to adults" range unclear **Population:** Hospital patients **Severity:** Unknown **History:** Unknown **Setting:** Unknown **Country:** Egypt	**GMP format:** Novel drink. Emulsion of corn oil and GMP in milk permeate. **Phe content:** Unknown. Finished Drink contained 12mg/ 100ml **LNAA content:** Unknown **Nutritionally complete:** Unknown. **Commercial product:** No. **% formula replaced:** Unknown. **Exchanges removed?** Unknown. **Same GMP in every patient:** yes **AA format:** Unknown. **Same AA in all subjects:** Unknown.	Phe measured at baseline and after 3 days (study-end). Conditions control data was collected under and method of analysis unclear. Method may be original Guthrie method which is only semi-quantitative.	(a) No control condition (as required in protocol). Control following three days of GMP compared to baseline. (b) Translated from Egyptian unclearly and vague, making it difficult to understand even basic aspects of the design. No response to a request for clarification was received. (c) Unrealistically large control improvements were reported (30- 80% over 3 days).

Table 8: Comparison of effect estimates and design quality: child studies.

Study	% of formula replaced by GMP foods. & sample size (n= #)	**Effect estimate.** (mean difference in plasma phe between GMP and formula conditions. Values are mean mean ± SE in µmol/L in fasting condition unless stated).	**Did control improve? (compared to formula)**	1. Random sequence generation (selection bias/ order effects	2. Concealment of allocations (selection bias/ order effects)	3. Blinding of participants & Personnel (performance bias)	4. Blinding of outcome Assessments (detection bias)	5. Incomplete outcome data (attrition bias)	6. Selective outcome Reporting (reporting bias)	7. Carryover in within-subjects Designs (carryover bias)	8. Influence of funding source (may introduce several biases)	9. Validity & reliability of outcome assessment (measurement error)	10. Reduced external validity	11. Sample size and statistical analyses (low power & precision)	12. Other threats to validity	Percentage of domains at low risk	% of important domains at low risk
----------Lower quality studies below. No higher quality studies were located----------																	
Daly et al., (2017)	Varied. 30 – 100% (n = 22)	GMP: 42 increase (sig) AA: 45 decrease (non sig) (Median change vs retro data) **GMP: sig higher (p <0.05)**	✘ Slightly Worse *	High Risk	High Risk	High Risk	High Risk	Low Risk	Uncl Risk	Low Risk	Low Risk	High Risk	High Risk	High Risk	Low Risk	**33**	**33**
Zaki et al., (2016)	50% (n = 10)	GMP: 376 (167 – 551) AA: 490 (289 – 597) Using: Median (interquartile) **No sig difference (p >0.05)**	✔ Equivalent	High Risk	High Risk	Uncl Risk	Uncl Risk	Uncl Risk	Uncl Risk	Low Risk	Low Risk	Uncl Risk	High Risk	Uncl Risk	Uncl Risk	**17**	**11**

*average difference between conditions is smaller than that observed by Zaki et al., (2016) but attained statistical significance due to larger sample and increased power.

Table 9: Compliance during GMP treatment in children.

Study	Compliance with protein substitute prescriptions (overall and frequency)	Compliance with other dietary requirements.	Where valid & reliable instruments used?	Was compliance clearly improved in the GMP condition?
Zaki et al., (2016)	**Unclear.** Clarification requested via email but no response received.	**Unclear.** Clarification requested but no response received.	Unclear. See previous columns.	**Unclear** Compliance not reported and clarification not received via email.
Daly et al., (2017)	**Unclear.** Compliance levels and frequency of intake are not reported and clarification was not received via email. The report mentions monitoring remaining stocks of protein substitutes during monthly home visits, but also identifies "failure to complete prescribed dose of protein substitute" as a possible confounder	**No significant difference between conditions..** ~50% more protein than recommended was consumed in both groups (5g extra in GMP, 4.5g extra in AA)	**Overall:** No (subjectively) **Protein subs:** Monthly home visits. **Diet:** 3-day weighed food diaries during the first and final week of each ~6 month condition & monthly home visits but authors note dietary non-compliance as a possible confounder	**Unclear.** No differences between conditions are reported. However, dietary compliance was monitored in detail only in the first and last weeks of ~6 month treatment periods and compliance levels with protein substitutes are not reported. The authors themselves identify non-compliance with both aspects as potential sources of confounding.

Table 10: Inter-individual variation in responses to GMP (both groups).

Group	Study	**Effect** (mean difference ± SEM between AA and GMP treatment) **and % of formula replaced by GMP.**	**Extent of inter-individual variation (**µmol/L**)**	**Where any determinants of inter-individual variation noted?**
Adults	van Calcar et al., (2009)	No significant difference AA: 619 ± 82 µmol/L GMP: 676 ± 92 µmol/L 100% replaced.	Responses varied from: 175 µmol/L decrease to 257 µmol/L increase	No. No consistent associations between the response to GMP, sex, genotype, and age were noted (using correlations)
Adults	Ney et al., (2016)	Significant increase of 147 ± 39 v AA (deterioration in control) AA: -285 ± 40 µmol/L (sig) GMP: +62 ± 40 µmol/L (non-sig) 100% replaced	Phe increased in 18 subjects & decreased in 12. Responses varied from: ~200 decrease to ~400 increase (using approximate 10^{th} & 90^{th} percentiles of box and whisker plot)	Yes. ANOVA noted interactions between control changes and: (a) total phe consumed through diet and GMP (b) Baseline phe concentration Suggesting these influence responses.
Children	Daly et al., (2017)	Significant increase of 87 µmol/L GMP: 42 increase (sig) AA: 45 decrease (non sig) 30 – 100% replaced	Variation in responses not noted as % replaced was adjusted for ethical reasons if control left safe ranges. Amount safely replaceable varied from 30- 100%	No. Authors suggest the amount of GMP introduced influenced responses & that tolerance influenced the amount safely replaceable but did not run correlations.

Table 11: Confidence in the body of evidence reviewed (GRADE score).

	Review of studies involving adults (16 years old and over)	**Review of studies involving children (3 – 16 years old)**
Patients	PKU patients (inpatients & outpatients)	PKU patients (inpatients & outpatients)
Outcome	Change in plasma phe control	Change in plasma phe control
Intervention	GMP based foods	GMP based foods
Comparator	Amino-acid formulas (usual treatments, which varied)	Amino-acid formula (usual treatments, which varied)
No of patients	63	32
Follow up	240 minutes to ~12 months	7 weeks to ~6 months
No of studies & designs	n=6: 2 randomised crossover trials (1 very short duration). 3 non-randomised studies. 1 self-controlled case report.	n=2: 2 Non-randomised pilot studies.
Initial score	**Low quality.** Evidence mainly from non-randomised studies (NRS).	**Low quality.** Evidence exclusively from non-randomised studies.
Risk of Bias	**Minus 1 point.** Several studies susceptible to bias (e.g. selective reporting) and/ or confounding (by diet/ compliance)	**Minus 1 point.** Both studies susceptible to bias (e.g. selective reporting) and/ or confounding (by diet/ compliance)
Inconsistency	**No deductions.** Results broadly consistent. Inconsistency apparently explained by bias/ design quality.	**No deductions.** Average amount of formula replaced identical (50%) but different approaches used.
Indirectness	**Minus 1 point.** One study selectively recruited patients liking the intervention and several used weak comparators.	**Minus 1 point.** The preliminary forms of GMP used differ from commercial products in organoleptic, LNAA and phe content.
Imprecision	**Minus 1 point.** CIs unavailable but 4 of 6 studies have too few subjects to provide adequate power so likely to lack precision.	**No deductions.** CIs unavailable but samples seem large enough to provide adequate power, suggesting acceptable precision.
Publication bias (meta-bias)	**Minus 1 point.** Evidence mainly from small, industry-linked, supportive, NRS. No small, non-supportive studies located.	**Minus 1 point.** Studies all NRS which face increased risk of publication bias (registrations unnecessary).
Quality of evidence base	**Very low (lowest possible score)**	**Very low (lowest possible score)**
Conclusion of review	Control is likely to deteriorate by ~10% on average after totally replacing formula with GMP but responses may vary. Patients may fail to compensate for the additional phe introduced by removing exchanges and control may deteriorate in some cases. Control and compliance should thus be monitored.	The very limited data available suggest GMP cannot totally replace formula in all children as in adults. The amount of formula safely replicable by GMP varies between individuals (~30- 100%) so if GMP is used, control must be monitored carefully and the dose adjusted if control leaves safe ranges.

4.0: Discussion

For complete details and evaluations of studies: tables 1, 3 & 5-8.

4.1: Objective 1a: effect of GMP on control in light of bias & design quality (adults).

Analyses contained two RS (Ney et al., 2016; Ahring et al., 2018) and four NRS (others in table 3). In such instances, Cochrane recommend separately analysing study types to avoid attaching undue weight to NRS (Reeves et al., 2011). The analysis instead separated studies by design quality (table 3) as one RS (Ahring et al., 2018) was at risk of inadequate randomisation and had other major limitations (below).

4.1.1: Objective 1a: Primary analysis (high quality studies).

Objective 1a was primarily addressed through one study (Ney et al., 2016). This randomised crossover compared 3 weeks treatment with 100% GMP and formula among 30 outpatients. Whilst its abstract reports equivalent control on GMP to baseline and the preceding narrative review accepted this, control was highly significantly worse than the formula comparator ($p< 0.0008$) by ~10% or +147 ± 39 µmol/L (mean ± SE). This change represents ~30% of ESPKU recommended ranges: 120 – 600 µmol/L (van Wegberg et al., 2017) suggesting it could compromise control in some cases. Furthermore, responses varied between individuals: control improved in 40% of subjects (by upto ~200 µmol/L) but worsened in 60% (by upto ~400 µmol/L). None-equivalent control may relate to compliance difficulties during GMP treatment as *every subject* failed to remove exchanges to compensate for GMPs phe content (as required) and an additional 88mg ± 6 phe was consumed. The self-reported dietary assessments used are error-prone but subjects consciously misreporting generally do so in desirable directions (Macdiardmid & Blundell et al., 1998; Prince et al., 1997) so patients would not be expected to so consistently misreport

overconsuming protein, suggesting reported intakes are accurate. An optimistic conclusion is that control would be better or even equivalent (on average) if compensation occurred. However, this was not demonstrated and difficulties compensating are concerning. Risk of most bias forms considered important in this context (table 2a) was low (table 3) and power was ample but one limitation may overstate GMPs benefit. The study exclusively recruited participants finding GMP acceptable, then permitted self-selection of GMP foods from a wide variety. This occurred before randomisation so threatens external (not internal) validity, something that depends on the scenario findings are applied to (CRD, 2009). However, most patients are unlikely to pre-sample GMP and select their preferred format from so wide a range, suggesting findings may overstate benefit when applied to the typical treatment scenario.

4.1.2: Objective 1a: Secondary analysis in (low quality studies)

Lower-quality studies generally reported equivalent control (table 3) but were susceptible to confounding and several bias forms likely to overstate GMPs benefit suggesting these more optimistic findings are misleading. Their findings were thus largely disregarded, though studies are evaluated briefly. In all five, randomisation was inadequate (Ahring et al., 2018) or absent (others). Prognostic imbalance is unlikely as all had self-controlled designs but preferential assignment to treatment periods could occur and usually overstates benefit (Higgins et al., 2011; Savovic et al., 2012). None used prospective protocols so selective outcome reporting may have occurred and generally overstates benefit (Dwan et al., 2013). Also, poor reporting left many design details unclear and clarification was only received for Ney et al., (2009) & van Calcar et al., (2009a). Given the low quality of studies, analyses considered poor reporting analogous to poor conduct, the most conservative approach.

Inpatient studies lack ecological validity as treatment is generally provided to outpatients and they artificially promote compliance. However, by better controlling for compliance differences they could directly describe GMP's impact on control. The two reviewed (Ahring et al., 2018 & van Calcar et al., 2009a) reported equivalent control alongside 100% compliance, supporting the above suggestion that broadly equivalent control would occur (on average) if compensation occured. However, in addition to the limitations noted above, both monitored control exclusively in the postprandial condition over unrealistically short timeframes (2 hours in Ahring et al., (2018) & 2 days in van Calcar et al., (2009a)) when treatment is lifelong and fasting control is equally important. Also, the comparison of interest in Ahring et al., (2018) (complete GMP formulation vs formula with an identical AA composition, minus phe) involved 6 patients so lacked power. Therefore, little confidence can be placed in these findings. Future inpatient studies should use randomisation, power calculations, prospective protocols, Consort guidance to clarify reporting, sufficient durations (at least three days per condition) and measure fasting and postprandial plasma phe.

The remaining three studies were even more limited (table 3). Only a case study (Ney et al., 2009) reported improved control (~10%) but this means little as the participant could be an outlier and the effect may be due to chance. The final two studies reported equivalent control but compliance formed an uncontrolled confounding variable as it was either not reported (LaClair et al., 2009) or monitored inadequately (Pinto et al., 2017). Pinto and colleagues also monitored control monthly, overlooking important fluctuations within months. The latter study was the only to partially replace formula in adults (27 - 100%). This reduces the need to remove dietary exchanges and may (theoretically) improve compliance/

control where this proves challenging but the study could not demonstrate the safety/ utility of the approach given its limitations.

4.1.3: Objective 1a: Summary & suggestions for research and practice (adults).

To minimise bias, conclusions are based on a single high-quality study (Ney et al., 2016). It selectively recruited patients finding GMP acceptable and allowed patients to select from many GMP foods, suggesting findings may prove optimistic in less ideal treatment scenarios. However, it was otherwise well-designed suggesting its findings (slightly inferior control on average following 100% replacement, partly relating to difficulties removing exchanges but inter-individual variation in responses) are valid. Better control would be expected if exchanges were removed to accommodate the phe in GMP. However, as only one high quality study was reviewed, these conclusions should be considered tentative until confirmed in similarly well-designed studies that better recreate practice by using less selective recruitment. Were several sufficiently strong, homogenous studies conducted, meta-analyses could confirm GMPs effect on control with increased power. However the lack of upcoming studies suggests this is unlikely, particularly as formula (which is universally recommended) is yet to amass such evidence (Yi, & Singh, 2015). Practioners prescribing GMP should expect slightly worse control on average but monitor phe closely as responses vary. As patients may struggle to remove exchanges, compliance should be monitored and dietary counselling offered if required. Partial replacement reduces the requirement to remove exchanges and could theoretically improve compliance/ control but the safety/ efficacy of the approach is yet to be demonstrated.

4.2: Objective 2a: Effect of GMP upon compliance (adults).

Compliance has three components. Firstly, consuming prescribed quantities of protein substitute (Macdonald et al., 2006; Duran, Rohr, Slonim, Guttler & Levy, 1999). Secondly, distributing this over 3-4 meals (Monch et al., 1996). Thirdly, consuming the correct quantities of dietary protein (and phe) and energy (to avoid catabolism which increases plasma phe). Regarding the quantity of protein substitute, no studies reported differences between conditions using self-reported logs (table 4). Ney et al., (2016) reported six subjects consuming insufficient GMP using objective plasma threonine measurements, raising concerns and suggesting protein logs overstate compliance, but *inferior* compliance was not objectively demonstrated as no objective measure of formula consumption exists. Regarding distribution of protein substitutes, the sole participant in Ney et al., (2009) reported improvements with GMP (three times per day vs once) citing its taste but may have been an outlier and the study used a weak comparator by comparing self-selected, ready-to-eat GMP to a powdered formula requiring preparation which may exaggerate compliance differences as formula is available in better-tasting, ready-to-drink formats (Macdonald et al., 2004; Prince et al., 1997). Ney et al., (2016) addressed this shortcoming in the strongest study and also reported improved distribution in the abstract but this is questionable as distribution only improved during a sub-analysis comparing patients receiving GMP and formula second. This distinction has no logical basis, the sub-analysis was not planned a-priori and distribution did not improve overall. Regarding dietary compliance, Ney and colleagues reported worse compliance on GMP as dietary protein was not removed (discussed previously). Overall, and contrary to expectations, the review found no reliable evidence of compliance improving on GMP-based therapeutic diets and some evidence that

removing exchanges may prove difficult. Future outpatient studies should clarify GMPs effect upon compliance using valid and reliable methods like the daily protein logs, plasma threonine monitoring and three-day weighed food diaries (per week) used by Ney et al., (2016). However, self-reported methods of assessing protein substitute (Prince et al., 1997) and dietary intake (Macdiardmid & Blundell et al., 1998) are error-prone, plasma threonine does not permit between-condition comparisons and inpatient studies risk introducing Hawthorne effects, so GMPs effect on compliance is difficult to describe accurately.

4.3.1: Objective 1b: effect of GMP on control in light of bias & design quality (Children).

As children require more protein (per kg BM) than adults for growth (WHO, 2007) but tolerate less phe (van Wegberg et al., 2017) formula comprises more of their diets. 100% replacement of formula with GMP thus appears impossible in some children as excessive phe is introduced and insufficient protein is consumed to remove in compensation. The two child studies reviewed (table 5) instead partially replaced formula with GMP. Zaki et al., (2016) compared treatment with 50% GMP, 50% formula to 100% formula and noted equivalent control as the 114 µmol/ L improvement in average control observed failed to obtain statistical significance. Daly et al., (2017) provided 12 children with GMP and a parrallel control group of 9 different children with 100% formula, monitored control for 6 months prospectively, compared this to 12 months historical control data (within-subjects) then compared changes between groups (between-subjects). Average control was significantly worse during GMP treatment, by 87 µmol/ L. However, both studies used different approaches. Zaki and colleagues replaced 50% of formula in all subjects whilst Daly and colleagues adjusted the proportion replaced on an individual basis if control left safe ranges for ethical reasons. The amount safely replaceable varied from 30 – 100% and whilst

it is likely similar inter-individual variation was noted by Zaki & colleagues (but manifested as variation in responses), individual patient data and variance data were not provided. What mediates this variation is unclear, though more GMP is likely to be replaceable in higher-tolerance patients. Findings suggest the amount of formula safely replaceable by GMP varies between children but this conclusion must considered extremely tentative as only two studies underpin it which have severe limitations.

One limitation is likely to understate GMPs benefit: the preliminary GMP formulations both studies used. Zaki & colleagues produced a spread from GMP much higher in phe (4.8 mg/ g Protein) than commercial products such as 'Glytactin' (1.8 mg/ g Protein) used by Ney et al., (2016), which must have impaired control. GMP lacks certain LNAA and requires supplementation to form a complete AA source. Daly and colleagues used GMP with an LNAA content meeting WHO (2007) requirements, less than Glytactin used by the Ney Group which contains 150% of the USA RDI for Tyrosine & 130% for other limiting LNAA (Van Calcar & Ney, 2012). This is likely to reduce the amount safely usable as LNAA appear to improve control by competitively inhibiting intestinal phe uptake (Matalon et al., 2006; 2007). Preliminary findings from a continuation of Daly's study (Daly et al., 2016) support this suggestion (table 6). Using the same design and subjects, the LNAA content of GMP was increased for five months and the average amount safely replaceable increased from 50% to 75%. This suggests more beneficial responses would occur with commercial GMP products lower in phe and higher in LNAA but this requires confirmation as the continuation of Daly & colleagues' study awaits publication so its details remain unclear and both studies have several limitations (below).

Neither published a protocol so selective outcome reporting, which generally overstates benefit, cannot be excluded (Dwan et al., 2013). Both lacked randomisation so selection bias cannot be excluded particularly in Daly et al., (2017) as concurrent groups selected treatments. Both studies also collected data under poorly controlled (clinical) conditions, risking confounding, particularly via differential compliance. Zaki & colleagues did not report monitoring compliance and Daly & colleagues identified compliance as a possible confounder since monitoring was inadequate (table 9). Both were also susceptible to measurement errors. Zaki and colleagues omitted important details regarding control measurements which affect findings such as time of collection, whether blood samples were fasting or postprandial and whether dried blood spots or blood draws were used. It is even unclear which analysis method was used and one of those mentioned appears semi-quantitative (nonderivatised amino acid/acyl-carnitine kits) calling the validity of findings into question. Daly and colleagues used DBS for between-subject comparisons. These recover 8-26% less phe than criterion methods but the error varies between individuals (Stroup et al, 2016) making such comparisons problematic. Further, between-subjects comparisons do not allow for differences in tolerance & baseline control, explaining why other studies use within-subjects designs.

4.3.2: Objective 1b: Summary & suggestions for research and practice (children).

Reviewed studies suggest the amount of formula safely replaceable by GMP without dangerous control loss varies between children, from ~30 – 100%. What mediates this variation is unclear though tolerance appears important (logically). As preliminary forms of GMP high in phe and low in LNAA were used in reviewed studies, more beneficial responses may occur with commercial GMP (e.g. Glytactin). However, conclusions are based on only

two studies susceptible to confounding, measurement errors and multiple forms of bias likely to overstate benefit (table 8) so conclusions should be considered extremely tentative until confirmed in stronger designs. These may be ethically required to adjust GMP doses individually like Daly and colleagues but should publish IPD and otherwise employ a randomised crossover design resembling Ney et al., (2016). This would prevent selection bias, permit within-subject comparisons and preclude the need for retrospective data. CONSORT compliance would ensure clear reporting and a prospective protocol would prevent selective reporting. Power calculations and oversampling would ensure adequate power and a validated plasma phe analysis method should be used, e.g. DBS analysed via tandem MS, which is convenient for outpatients as DBS collected at home can be posted. Ecological validity should be ensured by using GMP available in clinic. As residual phe content limits GMP's utility in children most, improved purification techniques should be investigated. Practioners choosing to prescribe GMP in children must be aware of the risks, monitor control carefully and adjust doses as the amount of formula safely replaceable varies. Whilst it is unclear what mediates this variation, less GMP is likely to be permissible in lower tolerance patients. GMP products high in LNAA and low in phe (e.g. Glytactin) should be selected as these appear to improve control and permit more replacement. Compliance should be monitored and counselling offered where needed.

4.4: Objective 2b: Effect of GMP upon compliance (children).

The review was unable to clarify GMPs effect upon compliance in children (table 9). Zaki & colleagues did not report monitoring compliance and Daly and colleagues identified non-compliance as an uncontrolled confounding variable because assessment methods were

inadequate. Future studies should rectify this using methods described for adults alongside the design features described previously.

4.5: Limitations of review

Inter-individual variation in responses occurred in adults (table 10) and in the amount of GMP permissible in children (30% - 100%) (Daly et al., 2017). However, the range of probable responses cannot be predicted with any confidence as only one study reporting variances was at low risk of bias (Ney et al., 2016). The review was also unable to identify mediators of this variation, which may predict patient responses. If several well-designed, homogenous studies published IPD, an IPD meta-analysis could quantify inter-individual variation and a meta-regression could identify its determinants (associatively) but this appears unlikely given the lack upcoming studies.

GRADE rates the confidence that can be placed in a review's conclusions by assessing the quality of the entire evidence base reviewed (Guyatt et al., 2011). Both reviews received the lowest possible rating (table 11) re-iterating that conclusions are extremely tentative. As the adult review was largely based upon a high quality study, this may appear overly critical but the decision to be conservative was considered safer.

GRADE also identified that both reviews were at risk of publication bias, despite the search strategy including grey literature and registries to prevent this. Language bias cannot be excluded as, without translation facilities, only English studies were reviewed. Publication bias cannot be excluded as NRS were included in both analyses and prospective registration of NRS is rare (Reeves et al., 2011). Both biases generally overstate benefit as supportive studies are disproportionately translated and published (Egger et al., 1997; Song et al.,

2000). Funnel plots and tests for publication bias were impossible as compatible precision data was unavailable due to poor reporting (CRD, 2009). Using a subjective approach, both analyses predominantly contained small, supportive studies with industrial links and no small, unsupportive studies, suggesting publication bias occurred (Guyatt et al., 2011). However such tests are unreliable, particularly in reviews containing few, similarly-sized, heterogeneous studies (Lau, Ioannidis, Terrin, Schmid & Olkin, 2006) so whilst publication bias appears likely, its presence cannot be confirmed. This may unfairly degrade the adult review that was primarily based upon a randomised, registered trial. However, registration of randomised trials is not mandatory so publication bias cannot be excluded. Indeed, one reviewed randomised study was unregistered (Ahring et al., 2018).

A PRISMA-P compliant protocol was prepared prospectively (appendix 3) but not registered, something not required by PRISMA-P (Moher et al., 2009). Resultantly, readers cannot exclude the possibility of bias in the review (e.g. changing analyses based on results).

Different results (statistical heterogeneity) are expected from studies with different designs/ populations (between-study heterogeneity). As statistical tests for heterogeneity have limitations, a subjective approach was used (Schünemann, Brożek, Guyatt & Oxman, 2013). Confidence intervals were unavailable due too poor reporting and whilst 'point estimates' varied both in magnitude (by 10-20%) and direction in adults, differences were (subjectively) attributed to design limitations (table 5). In children, statistical heterogenity was minimal as both studies reported (broadly) equivalent control but both used fundamentally different approaches making comparisons difficult. Given this, the subjective analyses used and the heterogeneous studies the relaxed inclusion criteria permitted, the contribution of

heterogeneity (e.g. differences in tolerance, history of control, setting and type of formula/ GMP used) to differences in effects cannot be excluded but it (subjectively) appears minimal as little statistical heterogeneity went unexplained.

Some adult studies contained patients under 16 (e.g. van Calcar et al., 2009a) but this is unlikely to affect results as examples were rare and all had tolerances comparable to adults. Moreover, all but one was over 12 and certain authorities (e.g. ESPKU) recommend relaxation of control at 12 rather than 16 (van Wegberg et al., 2017).

As the review addressed objectives previous reviews had not, findings cannot be compared to them. It is instead considered whether the systematic methods used addressed the reviews objectives and expanded upon conclusions of the preceding review. The exhaustive search methods located three additional adult studies: Laclair et al., (2009), Pinto et al., (2017) & Ahring et al., (2018) and the critical approach noted that Ney et al., (2016) reported a non-supportive findings, something the narrative review overlooked. In children a unique upcoming study was located (Daly et al., 2016) and an uncontrolled study (Abdel-Salam & Effat, 2010) reporting unrealistically optimistic control improvements that was previously reviewed but could have skewed results was excluded (table 7). The additional studies were bias-prone and contributed little to conclusions but this was not apparent beforehand; additional randomised trials may have existed. Moreover, authorities agree that identifying limitations in existing research and making recommendations for upcoming studies are themselves important functions of a systematic review (Schünemann et al., 2017; CRD, 2009; Moher et al., 2009), functions this review fulfilled.

The review included NRS, something Cochrane only recommend when randomised studies are impractical as misleading conclusions may be drawn due to bias (Reeves et al., 2011). In

adults the high quality study was analysed separately and conclusions based upon it, so misleading conclusions due to bias are unlikely. Given the few existing and upcoming studies involving children (none of which are randomised), the suggestion that children resist randomisation (Daly et al., 2017), wider difficulties facing PKU research including its low prevalence and difficulties applying adult findings to children, the inclusion of NRS involving children is defensible. It permitted important (albeit tentative) observations a review excluding NRS would overlook, namely that the amount of formula safely replaceable apparently varies between children and that the LNAA and phe content of GMP influence responses. Moreover, the same conclusion a review excluding NRS would draw was arrived at: that stronger studies are required. Given the transparent methods used and that the abstract stresses conclusions are tentative, conclusions are unlikely to mislead.

5.0: Conclusions

The review included six studies involving adults (over 16) but only a single randomised trial at low risk of bias (Ney et al., 2016). To minimise bias, conclusions were primarily based upon this. It suggests control will deteriorate by ~10% (or 147 µmol/L) on average if formula is totally replaced by GMP, which may prove clinically significant in some cases. However, responses varied from an ~200 µmol/ L improvement to an ~400 µmol/ L deterioration. Given this variation in responses, the mediators of which remain unidentified, monitoring control during treatment appears essential. The deterioration in control observed partly related to compliance difficulties as every subject failed to reduce dietary protein intake to accommodate the residual phenylalanine in GMP. Better control would be expected if dietary protein was removed but difficulties doing so could threaten control, so compliance should be monitored and counselling offered where required. The study had one limitation

that may overstate benefit: it selectively recruited patients finding GMP acceptable and permitted self-selection of GMP foods from a wide variety. Therefore, less beneficial responses may occur under less ideal treatment conditions. Contrary to expectations, the review found no reliable evidence of compliance improvements during GMP treatment and some evidence of compliance difficulties (aforementioned). However, as conclusions are based on a single randomised trial, they are tentative and caution is advised among practioners deciding to prescribe GMP to adults until further high quality studies are performed to clarify its effect upon control and compliance.

Two studies involving children (3 – 16) met inclusion criteria. These suggest GMP cannot totally replace formula in all children without plasma phe leaving safe ranges as tolerances are lower, formula comprises more of the diet and insufficient exchanges are consumed to remove in compensation. The amount safely replaceable appears to vary between children from ~30 – 100%. What mediates this variation is unclear though more GMP appears permissible in higher-tolerance children. However, both studies had heterogeneous and extremely limited designs in which selection bias, selective outcome reporting, measurement errors and uncontrolled confounding variables (particularly compliance) may have affected findings. Both also pilot-tested incomplete GMP formulations lower in LNAA and higher in phe than those used clinically, something likely to reduce effectiveness**.** Therefore, conclusions must be considered extremely tentative until confirmed in well-controlled randomised trials. Practioners choosing to prescribe GMP in children should exercise extreme caution, select products high in LNAA and low in Phe (e.g. Glytactin), monitor control and compliance carefully and reduce GMP doses if control leaves safe ranges.

6.0: Appendix 1: Full search strategy & results.

Table 12: Database-type resources A. databases & search engines containing published & unpublished literature

Resource searched	Date searched	Number of raw results	Resource type	Types of reports included	Approach used (no limits unless specified)
Pubmed	30/ 04/ 2018	502	Database	Published journals from MEDLINE including those awaiting indexing & online books	**Terms:** Phenylketonurias OR Phenylketonuria OR PKU OR "Phenylalanine hydroxylase deficiency" OR "PAH deficiency" OR hyperphenylalaninemia AND GMP OR Caseinomacropeptide OR Caseinomacropejohnptides OR Glycomacropeptide OR Glycomacropeptides OR "Protein substitute" OR "Protein substitutes" OR "amino acid formula" OR "phenylalnine free"
Cochrane Library (via search manager)	01/ 05/ 2018	386	Database	Published & unpublished RCTs and quasi RCTs from MEDLINE, EMBASE, registries, hand searches of key journals and a specialist IEM register. Cochrane & other reviews.	**Terms:** Phenylketonuria OR PKU OR "Phenylalanine hydroxylase deficiency" OR "PAH deficiency" OR hyperphenylalaninemia AND Caseinomacropeptide OR Glycomacropeptide OR GMP near/10 pku OR GMP near/10 phenylketonuria OR protein near/1 subset* OR "amino acid" near/1 formula OR "phenylalnine free" **Limiters:** (a) Cochrane reviews, other reviews, trials & Cochrane groups (other groups irrelevant) (b) "title, abstract & keywords"
CINAHL (via EBSCO host)	01/ 05/ 2018	280	Database	Published papers and magazines (that may include unpublished abstracts and other leads to unpublished papers). Specific to allied health professions (dietetics)	Phenylketonuria OR PKU OR "Phenylalanine hydroxylase deficiency" OR "PAH deficiency" OR hyperphenylalaninemia AND GMP OR Glycomacropeptide OR Caseinomacropeptide OR "Protein substitute" OR "Protein substitutes" OR "amino acid formula" **Expanders:** "apply related words"
Web of science (advanced search)	01/ 05/ 2018	234	Database	Published papers, books & conference proceedings (an important source of potentially unpublished studies). Wider coverage of subjects to increase sensitivity.	**Terms:** TS= (Phenylketonuria OR Phenylketonurias OR PKU OR Phenylalanine-hydroxylase-deficiency OR PAH-deficiency OR hyperphenylalaninemia) AND TS= (Caseinomacropeptide OR Caseinomacropeptides OR Glycomacropeptide OR Glycomacropeptides OR GMP OR CMP OR "Protein subst*" OR amino-acid NEAR/5 formula OR PKU NEAR/5 formula OR Phenylketonuria NEAR/5 formula OR phenylalnine-free)
Zetoc	02/ 05 /2018	72	Database	Published papers and conference proceedings (an important source of potentially unpublished papers).	**Approach:** The search facility has no support for Bolean operators and automatically links terms with AND. This was 'worked around' by separately searching all fields for: 1. "PKU AND GMP" 2. "Phenylketonuria AND GMP" 3. "PKU AND glycomacropeptide" 4. "Phenylketonuria AND Glycomacropeptide".

Resource searched	Date searched	Number of raw results	Resource type	Types of reports included	Approach used (no limits unless specified)
Google Scholar	03/ 05/ 2018	502	Internet search engine	Published papers, conference proceedings/ abstracts & dissertations/ theses. Crawls sites conventional databases may miss.	**Approach:** As search facilities lack advanced features searches generate many irrelevant results & these must be added to reference management software individually. To restrict searches & make results manageable, a single search was performed. **Terms:** "phenylketonuria AND glycomacropeptide"

Table 13: Database-type resources B. Databases & search engines specific to unpublished literature.

Resource searched	Date searched	Number of raw results	Resource type	Types of reports included	Approach used (no limits unless specified)
Open Grey (previously OpenSIGLE)	03/ 05/ 2018	23	European Grey literature database	Research reports, doctoral dissertations, conference papers, official papers & other grey literature.	**Terms:** Phenylketonuria OR PKU OR glycomacropeptide OR GMP NEAR PKU OR GMP NEAR phenylketonuria
New York Academy of Medicine	04/ 05 /2018	0	Grey literature specific database	Medical grey literature. Updates ceased in Jan 2017 but 1999 – 2017 is still searchable.	**Approach:** Used terms of increasing sensitivity. No results found for any permutation of either GMP or phenylketonuria.
Canadian Research Information system	04/ 05/ 2018	15	Database of research funding allocations	Studies that Canadian research funding was awarded to (potential source of published or unpublished literature)	**Approach:** Separate searches for PKU, Phenylketonuria & Glycomacropeptide.
Proquest Dissertations & Theses	04/ 05/ 2018	242	Database of dissertations & theses	Masters theses & doctoral dissertations	**Approach:** (Phenylketonuria OR PKU OR "Phenylalanine hydroxylase deficiency" OR "PAH deficiency" OR hyperphenylalaninemia) AND (Caseinomacropeptide OR Glycomacropeptide OR GMP near/10 pku OR GMP near/10 phenylketonuria OR protein near/1 subst* OR "amino acid" near/1 formula OR "phenylalnine free")
British Library EthOS	04/ 05/ 2018	33	Database of doctoral theses	Doctoral dissertations from UK institutions (not comprehensive)	**Approach:** Adavanced search limits searches to six terms. Separate searches were ran for "glycomacropeptide", "GMP" "phenylketonuria" & "PKU" to increase sensitivity.
Open Access Theses & Dissertations	04/ 05/ 2018	1	Dissertation & thesis search engine	Masters theses & doctoral dissertations.	**Approach:** advanced search permits only 4 terms: "phenylketonuria OR PKU AND GMP OR glycomacropeptide"

Table 14: Database-type resources C. Clinical trial & results registries.

Resource searched	Date searched	Number of raw results	Resource type	Types of reports included	Approach used (no limits unless specified)
International clinical trials registry platform (ICTRP)	03/ 05/ 2018	127 trials	Database of clinical trial registries	World Health Organisation database completely covering most major regional/ national trial registries. There is & 1- 4 week lag when indexing registries but trials so recent will not have reported findings.	**Terms:** Phenylketonuria OR PKU OR PAH deficiency OR hyperphenylalaninemia AND GMP OR Glycomacropeptide OR Caseinomacropeptide
CenterWatch Clinical Trials Listing Service	03/ 05/ 2018	7	Commercial database of clinical trial protocols	Clinical trial protocols	Only searches by condition permitted. Searched for "PKU" & "phenylketonuria".
Registries maintained by pharmaceutical companies	03/ 05/ 2018	3	Trial protocol/ results registries	Trial protocols and results which companies may only publish on their own registries.	**Approach:** Separate searches for "PKU, phenylketonuria, GMP & glycomacopeptide" **Companies included:** Astrazeneca, Bayer, GSK, Eli Lilly, Bristol-Myers Squibb & Novartis. Most now register all trials on clinicaltrials.gov & its European equivalent and no longer maintain databases.
clinicaltrial results.org	03/ 05/ 2018	0	Website containing trial results	Clinical trial results	**Approach:** Separate searches for "PKU, phenylketonuria, GMP & glycomacopeptide"

Table 15: Snowball-type resources A. Reference list searches

NB: results from these resources were screened as searches were conducted.

Resource searched	**Date searched**	**Unique, screened reports**	**Description of resource**	**Approach used (no limits unless specified)**
Reference lists of reports screened for inclusion	16/ 05/ 2018	0	Searches of reference lists from studies that passed screening.	**Search:** Searched manually & using 'ctrl + F' searches for "macro", "glyco" & "GMP". Unique results then screened for inclusion. **Included studies:** Abdel-Salam & Effat, (2010), Daly et al., (2017), Zaki et al., (2016), Ney et al., (2009), van Calcar et al., (2009a), LaClair et al., (2009), Ney et al., (2016), Pinto et al., (2017), Ahring et al., (2018).
Reference lists of reviews	17/ 05/ 2018	0	Searches of reference lists from reviews located via database-type resources and during the preceeding narrative review, for eligible studies	**Search:** Searched manually & using 'ctrl + F' searches for "macro", "glyco" & "GMP". Unique results then screened for inclusion. **Included reviews:** Blau et al. (2010) Harding & Blau (2010), Macleod & Ney (2010), Poustie & Wildgoose (2010), van Spronsen & Enns (2010), Blau, Hennerman, Langenbeck & Licther-Konecki (2011), Giovannini, Verduci, Salvatici, Paci & Riva (2012), van Calcar & Ney (2012), Ho & Christodoulou (2014), Ney et al., (2014), Strisciuglio & Concolino (2014), Al Hafid & Christodoulou (2015), Blau & Longo (2015), Yi & Singh (2015), Rocha & MacDonald 2016, Ney & Etzel (2017), Spécola & Chiesa (2017)
Reference lists of clinical guidelines	17/ 05/ 2018	0	Searches of reference lists from clinical guidelines located via database-type resources and during the preceeding narrative review, for eligible studies	**Search:** Searched manually & using 'ctrl + F' searches for "macro", "glyco" & "GMP". Unique results then screened for inclusion. **UK guidelines:** Smith et al, (1993) **German guidelines:** Burgard et al., (1999), **EU guidelines:** van Spronsen et al., (2017), van Wegberg et al, (2017), **USA Guidelines:** NIH CDP (2001), Vockley et al., (2014)

Table 16: Snowball-type resources B. Citation searches

Resource searched	Date searched	Unique, screened reports	Description of resource	Approach used (no limits unless specified)
Citation searches of reports screened for inclusion	17/ 05/ 2018	0	Takes reports, & doctoral dissertations (located via database-type resources) that passed screening and locates articles that have cited them which may be eligible for inclusion	**Search:** Title of paper inserted into (a) Web Of Science (b) Google Scholar. Unique results (not already located) then screened for inclusion. **Included Reports:** See previous table **Included Dissertations:** See previous table
Citation searches of reviews	17/ 05/ 2018	0	Takes reviews located via database-type resources and the preceeding narrative review and locates articles that have cited them which may be eligible for inclusion	**Search:** Title of paper inserted into (a) Web Of Science (b) Google Scholar. Unique results (not already located) then screened for inclusion. **Included reviews:** See previous table
Citation searches of clinical guidelines	17/ 05/ 2018	0	Takes guidelines located via database-type resources and the preceeding narrative review and locates articles that have cited them which may be eligible for inclusion	**Search:** Title of paper inserted into (a) Web Of Science (b) Google Scholar. Unique results (not already located) then screened for inclusion. **Included guidelines:** See previous table

Table 17: Snowball-type resources C. Handsearches of key journals

Resource searched	Date searched	Unique, screened reports	Rationale for handsearching	Approach used (no limits unless specified)
Molecular genetics & metabolism (journal)	18/ 05/ 2018	2	Contained several relevant reports, particularly conference proceedings from the Annual Meeting of the Society for Inherited Metabolic Disorders (SIMD) and the International Congress of Inborn Errors of Metabolism (SSIEM)	(a) Web-based journal issues from 2007 (when GMP was first used in PKU) were searched for relevant articles using titles and abstracts, including in-press articles (b) Conference proceedings were searched manually and using 'ctrl + f' searches for "GMP", "glyco" and "macro" (c) potentially relevant, unique papers were screened for inclusion

Resource searched	Date searched	Unique, screened reports	Rationale for handsearching	Approach used (no limits unless specified)
Journal of inherited metabolic disease	18/ 05/ 2018	1	Contained several relevant reports & abstracts from the Annual Symposium of the Society for the Study of Inborn Errors of Metabolism	As above
American journal of clinical nutrition	18/ 05/ 2018	0	Contained several relevant reports	As above

Table 18: Snowball-type resources D. Manual searches of key websites

Resource searched	Date searched	Unique, screened reports	Resource type	Types of reports included	Approach used (no limits unless specified)
Websites of manufacturers of PKU formulas & treatments	14/ 05/ 2018	0	Company websites	Published & unpublished reports of studies supporting products	**Approach:** Site search facilities were used to search for "PKU, phenylketonuria, GMP & glycomacopeptide" & entire sites were handsearched using sitemaps. Potential reports were screened. **Companies included:** Biomarin, Cambrooke, Vitaflo UK, Vitaflo USA, Nutricia UK, Mead Johnson & Abbot Nutrition.
Website of European PKU society (ESPKU)	14/ 05/ 2018	0	Website of PKU Society	Published papers & abstracts/ proceedings from international conferences.	**Approach:** Site searched as above. Conference proceedings from 2003 & 2012 were located on the NSPKU site (below) & reports screened for inclusion. Email correspondence indicates the content of other years conferences are only available to attendees.
Website of English National PKU Society (NSPKU)	14/ 05/ 2018	0	Website of PKU Society	Published papers & abstracts/ proceedings from national conferences.	**Approach:** Site searched as above. Conference proceedings from 1998- 2002, 2005- 2007 & 2016- 2018 were located, handsearched & screened. Others were requested via email, but no response was received after 7 days.
Website of US National PKU Alliance	14/ 05/ 2018	0	Website of PKU association	Conference presentations & funding awards which may identify studies.	**Approach:** Site searched as above. Research projects funded by NSPKU & conference presentations from 2010 onwards were handsearched & screened for inclusion.

Resource searched	Date searched	Unique, screened reports	Resource type	Types of reports included	Approach used (no limits unless specified)
Website of canpku (Canadian PKU Association)	14/ 05/ 2018	0	Website of PKU association	Lists published studies & minutes from events.	**Approach:** Site searched as above. Listed studies were screened for inclusion.
Website of Irish PKU Association (pku.ire)	14/ 05/ 2018	0	Website of PKU association	None	**Approach:** Site searched as above.
Website of New South Wales PKU Association	14/ 05/ 2018	0	Website of PKU association	None	**Approach:** Site searched as above.
Website of British Dietetic Association	14/ 05/ 2018	0	Website of dietetic association	Abstracts from annual research symposiums, published in the journal of human nutrition & dietetics.	**Approach:** Site searched as above. Most content requires membership. Symposium proceedings were available in issues of the journal of human nutrition & dietetics. Those dating back to 2008 were handsearched, and relevant reports screened for inclusion.
USA Academy of Nutrition & Dietetics site	14/ 05/ 2018	0	Website of dietetic association	Maintain a database of reviews & guidelines & the Journal of the Academy of Nutrition and Dietetics	**Approach:** Site searched as above. The database is inaccessible to non-members & the journal is indexed by PubMed (searched above)
Website of dietitians of Canada	14/ 05/ 2018	0	Website of dietetic association	Abstracts from annual conferences, published in the Canadian Journal of Dietetic Practice & Research.	**Approach:** Site searched as above. Some content is inaccessible to non-members. Symposium proceedings are available in the Canadian Journal of Dietetic Practice & Research. Proceedings from 2008 were handsearched, and relevant reports screened for inclusion.
Dietitians Association of Australia website	14/ 05/ 2018	0	Website of dietetic association	Conducts annual conferences & produces a journal.	**Approach:** Site searched as above. Conference abstracts were not available and were requested via email but no response was received within the predetermined seven day window.
Confederation of Dietetic Associations Website	14/ 05/ 2018	1	Website of international dietetics confederation	Conducts the international congress of dietetics (an international conference) every four years	**Approach:** Site searched as above. Abstracts from the 2008, 2012 & 2016 international congress conferences were handsearched. Others predate the first use of GMP in PKU (~2007) and the next is in 2020.
Nutrition Society website (international)	14/ 05/ 2018	0	Website of nutrition society	Conducts conferences & publishes six journals, one of which contains abstracts from these conferences.	**Approach:** Site searched as above. All six journals are indexed by pubmed (searched previously). Conference abstracts were searched as above.

6.1: Appendix 2: Example forms.

Table 19: Blank screening Checklist

Screening Criteria	**Yes, No, or Unsure**	**Notes.**
Is the study in English?		
Are the participants human?		
Are the participants diagnosed with PKU?		
Does the intervention replace any amount of the participants usual amino acid formula with any amount and type of GMP?		
Are phenylalanine levels monitored?		
Does the study involve a control group?		
Does the study meet screening criteria?		

Instructions: Only reject reports clearly failing to meet one criteria. If unsure include the study but note the source of uncertainty.

Table 20: Blank inclusion criteria checklist

	Screening Criteria and explanation.	**Yes, no, unsure**	**Notes.**
	Are the participants diagnosed with 'conventional' forms of PKU? --Studies containing subjects with BH_4 defects (that require differential treatment) and maternal PKU (which reduces tolerance and complicates management considerably) are excluded.		
	Where the patients treated during childhood (either at birth or late treated) and did they require dietary treatment to maintain control at the onset of the study? (whether currently receiving treatment or not)? --Studies involving untreated patients are excluded as these are unrepresentative of the wider PKU population and may have severe neurocognitive disabilities impairing their ability to adhere to the protocol. --Patients with mild forms of PKU not requiring treatment to remain within recommended control ranges are also excluded. 'Patients requiring treatment' is specified for inclusion instead of severity thresholds as these are applied inconsistently (e.g. between countries) and the assessment of severity is problematic (its is determined before treatment at birth, often before blood phe peaks). Studies involving untreated or milder cases that provide separate data treated patients with 'conventional/ classical PKU' are permitted.		
	Are the patients free of other metabolic, liver and kidney conditions and other conditions that may affect protein metabolism?		
	Were patients receiving other treatments such as LNAA or Sapropterin? If patients are receiving sapropterin (synthetic BH_4) co-treatment, was their phe tolerance stable before the study and was the sapropterin dose consistent throughout the study? --Exclude studies involving patients taking LNAA as their effects are not well characterised and may confound those from GMP in unpredictable ways. --Sapropterin (synthetic BH4) treatment is common in the USA and excluding patients receiving it would make recruitment difficult and reduce ecological validity but it has the potential to confound effects otherwise attributable to GMP if not used consistently throughout.		
	Is blood or plasma phe monitored at baseline and on at least one other occasion during the study, for both conditions?		
	Is the full text of the study available in English?		Record foreign otherwise eligible.
	Does the report meet inclusion criteria?		

Instructions: If unsure, describe the reason for this uncertainty in the notes column and continue to assess the report. The author will be contacted for clarification

Table 21: Example data extraction form

Study Assessed	**Ney et al., (2016). Glycomacropeptide for nutritional management of phenylketonuria: a randomized, controlled, crossover trial**
Summary and Important Notes (complete after data extraction):	**Summary:** Randomised crossover trial comparing control (using fasting measurements) in n=30 free living patients (mainly adults aged 15-49y) during 3 weeks usual treatment (formula) to 3 weeks GMP treatment, separated by a 3 week washout on formula. There was a significant difference in control between conditions (using group means) due to a small but significant improvement on formula and a small, non-significant deterioration on GMP. However, whilst control was not totally equivalent between treatments: (a) the difference between conditions is of borderline clinical significance (b) the small loss of control noted on average for GMP (vs baseline) was not statistically or clinically significant (c) Responses varied individually: around 40% of patients maintained or improved control on GMP (d) Dietary phe intake increased significantly ($p = 0.0259$) in the GMP condition (by 88 ± 6 mg Phe/d) despite the diets being designed to contain equal phe content, because subjects did not remove as much dietary phe (exchanges) as directed. Had exchanges been removed as directed, the difference between conditions and the extent of deterioration of control on GMP would have been even smaller and less concerning. However, it could be argued that this study provides evidence that (a) control can be modestly worse on GMP in some individuals (b) that the requirement to remove exchanges after GMP substitution is challenging and may promote control loss in some individuals. Careful monitoring and additional support/ education may be warranted. Ideally, the phe content would be reduced so fewer exchanges need to be removed and newer products are emerging with lower phe content.
Links to supplemental tables, errata, claims of fraud etc.	None

Factor	**Explanation of factor and information required**	**Description of factor as stated in study and location (e.g. methods)**	**Notes or missing information that must be requested from authors (note what is required)**
Date of form completion		01-06-2018	
Author contact details (ideally an email address)	To whom correspondence should be addressed.	Professor Denise Ney ney@nutrisci.wisc.edu	
Publication type/ status	e.g. published, abstract, protocol	Published Paper	
Unit of allocation	e.g. individuals or cluster	Individuals	
Study design (archetype)	e.g. RCT, quasi RCT, crossover, NRS, other design (specify)	Randomised crossover trial	
Detailed overview of study design	Describe the overall design including: duration, whether parallel or serial, which aspects are prospective	Randomised crossover trial comparing control (fasting measurements) in n=30 free living patients (adults aged 15-49y) during 3 weeks usual treatment (formula) to 3 weeks GMP treatment, separated by a 3 week washout on formula.	

Factor	Explanation of factor and information required	Description of factor as stated in study and location (e.g. methods)	Notes or missing information that must be requested from authors (note what is required)
Number of participants		30	
Age of participants and baseline imbalances in parallel NRS.	Range and average. Studies are likely to only involve one group, or perform separate analyses on each. If they do not note imbalances as they may confound effects.	15-49	
Do ages fit this reviews age brackets (<16 & 16 and over)>	Note exceptions and consider in discussion/ appendices.	No: Age range is 15-49 but only 5 patients are under 17.	Grouping this study with other adult studies in unlikely to negatively affect the review as: (a) only 5 patients are under 17 (b) The tolerance and energy requirements of 15 year olds are similar to adults (unlike younger children) (c) The tolerance of all included patients were comparable and 100% replacement was performed in them all. The study was therefore grouped with other adult studies in this review.
Population participants are drawn from and sampling design.	E.g. university students, hospital patients, care homes? See quality assessment pt. 13	PKU patients recruited from: (a) Biochemical Genetics Program, Waisman Center, University of Wisconsin-Madison (b) the Division of Genetics and Genomics, Boston Children's Hospital, (c) Harvard Medical School (d) National advertisement within the phenylketonuria community	
Setting & staffing arrangements	e.g. inpatient, outpatient, university etc.	Outpatient: patients were at home but returned to centres for blood phe measurements.	
Country		USA	
Number of centres (for study conduct)		2 centres Waisman centre (n=19) Boston Children's Hospital (n=11)	

Factor	Explanation of factor and information required	Description of factor as stated in study and location (e.g. methods)	Notes or missing information that must be requested from authors (note what is required)
Tolerances and baseline distribution between groups in parallel NRS.	In mg phe (ideally) or g protein. Note distribution imbalances and countermeasures in parallel NRS See quality assessment tool pt 1	Severity and baseline control data cover this	
Baseline Protein substitute prescriptions (g/ PE) and distribution between groups at baseline in parallel NRS.	Reflects tolerance as treatment is individualised by tolerance. Therefore unnecessary if tolerance is reported. Note distribution imbalances. See quality assessment tool pt. 1	Severity and baseline control data cover this	
Severity of condition and baseline distribution between groups in NRS (only to be used **if previous two factors are unavailable**).	Does not reliably predict tolerance/ substitute requirements. Note distribution imbalances and measures. See quality assessment tool pt 1	Severities: 20 classical, 10 variant (based on genotype and response to synthetic BH_4).	**See:** (a)Characteristics of participants section Note point below
Baseline control levels/ history of control and distribution between groups in parallel NRS	Potential confounder in NRS. See quality assessment tool pt 1	**Baseline control levels:** baseline plasma phe was significantly higher among patients with classical PKU (867 ± 73 mmol/L) than those with variant PKU (461 ± 59 mmol/L $P < 0.001$ **History of control:** all subject early treated (shortly after birth). Optimal control was not required.	Baseline control varied with severity and the initial ANOVA performed analysis suggested baseline control interacted with the main effect (I.E. the change between plasma phe at baseline and the end of each treatment period was different for patients with worse severities and higher baseline plasma phe). However, a pre-planned ANCOVA was performed that incorporated this potential source of confounding into analyses, so it is not problematic.
Any important co-morbidities, and baseline distribution in parallel NRS	E.g. liver or kidney disease affecting protein metabolism. Note imbalances. See quality assessment tool pt 1	None	

Factor	Explanation of factor and information required	Description of factor as stated in study and location (e.g. methods)	Notes or missing information that must be requested from authors (note what is required)
Type of formula administered (comparator)	Format? (powder or modern?) Does it vary inter-individually? See quality assessment pt 12.	**Format:** Mixture of modern and traditional (some patients consumed more than one type): 21 conventional powder in can 4 powder in can or pre-portioned packets 14 liquid ready to drink liquids 2 pills **Differences between subjects:** yes, usual treatment involving 15 different products	Many subjects consumed modern versions of formula so this study is a more fair and ecologically valid comparison than some others
Type of GMP consumed (intervention)	Format? (powder or modern?) Nutritionally complete? Used clinically or prelim form? Does it vary inter-individually? What is Phe content? See quality assessment pt 12.	**Formats:** self-selected modern foods (pre-portioned drink mixes, pre-portioned pudding mixes, sports drinks) LNAA content: High vs formula (150% of USA RDI for Tyr, 130% for other limiting LNAA) **Complete:** yes **Differences between subjects:** yes, self-selected **Phe content:** 1.8mg/ g PE (comprised of 70% cGMP-20 from Arla Foods Ingredients and 30% supplemental AAs)	Subjects selected their favourite GMP foods during a screening period. This is addressed below as a possible source of bias **Notes:** (a) LNAA: His, Leu, Met, Trp, Tyr, Thr, Val, Ile, Phe. (b) GMP naturally high in: Thr, Ile. (b) Supplementation needed for: Arg, His, Leu, Trp, Tyr & Micronutrients.
Proportion of formula replaced by GMP and dietary adjustments made?	E.g. Replaced 100% and removed dietary exchanges to completely adjust.	100% replacement. Dietary exchanges were supposed to be removed so diets contained equal phe between conditions	
Was dosing individually adjusted for tolerance/ severity or sub-analyses performed?	Flag for separate analysis for objective regarding individual variation in responses. Detail dosing strategy (tables/ figs)	No: 100% replacement in all cases	

Factor	Explanation of factor and information required	Description of factor as stated in study and location (e.g. methods)	Notes or missing information that must be requested from authors (note what is required)
Compliance levels with protein substitutes	Total compliance and frequency of ingestion. Note differences between groups See quality assessment point 10	**Overall:** Compliance was not significantly different between conditions (p = 0.576) and was considered "adequate" **Prescribed Protein sub intake:** 0.85 ± 0.03g PE / kg BM / d **Compliance with GMP via diaries:** 0.74 ± 0.04 g PE / kg BM / d **Compliance with formula via diaries:** 0.76 ± 0.05 g PE / kg BM / d **Compliance with GMP using plasma Thr levels:** 6 of 30 subjects displayed reduced plasma Thr between baseline and the end of GMP treatment, suggesting poor compliance in the 72hr before the blood draw. Sub-analysis excluding them did not alter study findings. **Frequency:** GMP was consumed significantly more frequently than AA (3.74 ± 0.24 Servings/d vs 2.43 ± 0.24 Servings/d, p = 0.001), **BUT** only when comparing GMP and AA consumption during visits 3-4 (interaction with sequence noted via ANOVA, that was not specified a-priori in protocol or expected)	**(b) greater frequency of consumption with GMP was not noted overall:** (only via an interaction with time in the ANOVA which was unexpected I.E. when comparing the 15 patients receiving GMP last to the 15 receiving GMP last) This was clarified via email.
Method of ensuring and assessing compliance with protein substitutes	Method, timeframe and frequency of monitoring. See quality assessment point 10.	(1) Total Intake and frequency of intake of GMP/ formula monitored through daily medical food logs (self-reports) (2) GMP intake objectively monitored using plasmas threonine. GMP is high in Thr and known to elevate plasma Thr. Reduced plasma Thr after stabilisation indicates poor compliance.	

Factor	Explanation of factor and information required	Description of factor as stated in study and location (e.g. methods)	Notes or missing information that must be requested from authors (note what is required)
Compliance levels with dietary advice	Particularly protein and energy intakes (total, frequency). Note differences between groups. See quality assessment point 11	Dietary phe intake increased significantly ($p = 0.0259$) in the GMP condition (by 88 ± 6 mg Phe/d) despite the diets being designed contain equal phe content, because subjects did not remove dietary phe (exchanges) as directed in the GMP condition. Intake of Phe from natural foods was not different between conditions and did not change significantly compared with baseline for either treatment.	
Method of assessing dietary compliance	Method, timeframe and frequency of monitoring. See quality assessment point 11	(a) 3 day food records completed before each study visit (baseline, end of AA, end of washout, end of GMP) Email correspondence suggests "Food diaries were largely estimated based on patient experience using household measures; a few patients did weigh their food with gram scale." (b) Nutrient content estimated by skilled dietitian using Food Processor SQL and checked by second skilled assessor. Phe content of some foods incorporated using manufacturers information and Virginia Schuett's Low Protein Food List for PKU	(a) food diaries are error prone (they were mainly estimated though some were weighted) (b) Food diaries only covered the final 3 days of each 3 week treatment period, these 3 days may not be typical
Administration of co-interventions, noting imbalances	e.g. BH_4 or education. Note differential administration between groups.	Patients receiving Sapropterin were eligible provided (a) their tolerance was stable (c) the dose of Sapropterin remained constant over the study.	Note this and consider impact in quality assessment
Duration of intervention periods, noting imbalances		3 weeks each, no imbalances	

Factor	Explanation of factor and information required	Description of factor as stated in study and location (e.g. methods)	Notes or missing information that must be requested from authors (note what is required)
Time points measurements were taken at (duration, frequency, time of day)	Also see Qual assessment point 9	(a) 4 fasting venous blood samples taken during study visits (1) baseline pre-AA (2) end of AA (3) baseline pre-GMP (4) end of gmp (b) dried blood spots taken at 8 points during each 3 week treatment period (on days 3, 6, 7, 10, 13, 14, 17, 21)	
Time points measurements were reported for	Note discrepancies with above	All reported	
How samples were collected	DBS or blood? Plasma or serum? See Quall assessment point 9	(a) Analysis of plasma AA concentration from venous blood samples using Hitachi L-8900 amino acid analyser (with post–column ninhydrin derivatization ref 30 therein). Interassay CV = 1.6% (b) Analysis of plasma phe from dried blood spots using tandem MS and duplicate samples (nonderivatized flow-injection method used for newborn screening). Interassy CV = 11.7%.	
How samples were analysed **and validity/ reliability** data for method	E.g. MS-MS or Guthrie assays. See Quall assessment point 9		
Any differences between conditions in assessments?	Potential to confound	No.	
Were any sub-analyses performed, and were these selected a-priori?	Report the analyses, findings and whether predetermined or post-hoc.	In the ANOVA, an interaction was noted between the main effect (a comparison of the change in plasma phe from baseline between each condition) and 2 factors: (a) Baseline control level I.E. baseline blood phe – which varied with severity (being higher where severity was worse) (b) Dietary phe intake This was adjusted for by importing these factors as co-variates in a pre-planned ANCOVA.	**NB:** some limited evidence of factors that predict inter-individual variation in responses to GMP: (a) baseline control level (baseline plasma phe) (b) severity (subgroup analyses suggested smaller, non-sig changes in control in milder variant forms – but lacked power) (c) dietary phe intake I.E. subjects with worse baseline control, severity and greater dietary phe intake saw greater changes in plasma phe so may be at greater risk of control loss.

Factor	Explanation of factor and information required	Description of factor as stated in study and location (e.g. methods)	Notes or missing information that must be requested from authors (note what is required)
Effect estimate, measure of variability and analysis used	e.g. mean difference & standard deviation/ standard error/ 95% confidence intervals e.g. ANOVA, t-test	**Results (a)** (a1) There was a significant difference between conditions: (-147 ± 39 µmol/L , p = 0.0008) using mean ± SEM difference (a2) GMP-MFs showed a non-significant increase in plasma phe 62 ± 40 µmol/L, P = 0.136) (mean ± SEM) (a3) AA-MFs showed a modest, significant decrease in plasma Phe concentration (285 ± 40 µmol/L, P = 0.044) (mean ± SEM) **Analysis (a) using venous blood draws:** Comparison of change in plasma phe from baseline to period-end (3 weeks) between conditions was investigated using ANCOVA to adjust for differences in baseline plasma phe and dietary phe intake (covariates). See next column for details. **Results (b)** The proportion of subjects showing an increase in plasma phe when comparing GMP to AA was not significantly different (p=0.267): 60% increased on GMP, 40% decreased on GMP. **Analysis (b) using venous blood draws:** Proportion of subjects whose plasma phe rose between baseline and study end compared between conditions using Mcnemar's Test.	**Analysis (a) :** ANOVA conducted initially to compare changes in plasma phe from baseline to period-end between conditions (the main effect) and check for interactions with other variables. Interactions between the main effect and (a) baseline plasma phe (b) 'dietary phe intake' were noted, so ANCOVA was performed, incorporating these factors as covariates.

Factor	Explanation of factor and information required	Description of factor as stated in study and location (e.g. methods)	Notes or missing information that must be requested from authors (note what is required)
Statistical significance reported and significance level used	e.g. p = 0.02 at significance level of 0.05	See above (a = 0.05)	
Sample size for effect, if different	See quality assessment pt. 14	Same	
Was power calculation performed?	See quality assessment pt. 14	Yes: "The trial was powered at 80% (b = 0.20) based on a previously reported SD of 150 mmol/L for plasma Phe (20, 32) to detect a change in plasma Phe of 120 mmol/L at a = 0.05 (2-sided test). This provided a sample size of 25 subjects. 30 subjects were recruited with an estimated dropout rate of 15-20%.	
Did study have adequate power and precision?	See quality assessment pt. 14	**Power:** yes, see above **Precision:** yes	
Overall appropriateness of analysis and impact on findings/ weighting	See quality assessment pt. 14	Analyses appear suitable	

Table 22: Example data request form

Data request form: Ney et al., (2016)		
Factor information is required for	**Example of type of information required**	**Response from you RE information, with explanation where necessary**
Report that 'frequency of consumption' improved in GMP condition vs formula condition.	Greater frequency of consumption was noted during the GMP condition, via an interaction with sequence in the ANOVA. It is unclear (to me) what is being compared here. Is this improved frequency of intake when comparing n=15 patients receiving GMP last to n=15 patients receiving AA last? --Can you please (a) clarify this (b) confirm that no interaction with sequence was expected (b) confirm that this analysis was not specified a-priori?	Yes, your interpretation is correct; it made stats for subjects during their 2nd or last of the two treatments. We did not expect that the order of treatments would affect the results, but tested for a sequence effect in all analyses. Analysis for frequency of MF intake was included in the approved protocol before subject recruitment began; it was a secondary outcome.

Factor information is required for	**Example of type of information required**	**Response from you RE information, with explanation where necessary**
Method of assessing dietary compliance	Were food diaries weighed or estimated, and what method was used for weighing/estimates?	Food diaries were largely estimated based on patient experience using household measures; a few patients did weigh their food with gram scale.
Random sequence generation	Randomisation of treatment order was employed.. "Equal randomization of the diet treatment order was achieved by using a computer-generated scheme." However, 10 subjects that found GMP unacceptable during a screening period were not enrolled. I assume randomisation occur after this screening period?	Yes, if GMP unacceptable then not enrolled. Yes, randomization occurred after screening and consenting – at the time of enrolment.
Concealment of allocation	Were upcoming allocation concealed? What method was used?	No subjects and research staff knew the allocation so they could counsel subjects. **NB:** misinterpreted as 'unblinding'. Randomisation confirmed as adequate via subsequent email.
Blinding of participants and personnel	Blinding of participants? (yes/ no) Blinding of personnel? (yes/ no) Method?	No blinding, impossible as products taste and look very different.
Blinding of outcome assessment	Blinding? Method?	"
Incomplete outcome data	I note two subjects dropped out because they were unable to complete requirements of the protocol because of time constraints How much data was obtained for them, and how was their dropping out handled?	Both dropped out within 1st week of first treatment. Very little data was obtained other than baseline and no data was included in analysis from these 2 subjects.

Table 22: Example quality assessment form:

Quality assessment form: Ney et al., (2016)			
Design feature (and number for guidance notes)	**Risk of bias, confounding, imprecision or poor external validity. (unclear, low, high or n/a)**	**Explanation of decision, with explicit reference to methods used and measures taken to minimise impact.**	**Where risk is high or unclear, how important is this source of bias/ confounding and what is the likely direction and magnitude of its impact.**
Random sequence generation (1)	Low risk of bias (from order of treatments or trends over time)	The order of treatments was generated using computer-based randomisation considered adequate by Cochrane guidance and ANOVA found no evidence of sequence effects so randomisation was adequate.	**NB.** Subjects sampled GMP and those finding it unacceptable were excluded from the study but this occurred before randomisation at the sampling stage. It is therefore considered a threat to external validity in point 10.
Concealment of allocation (2)	Low risk of bias (from order of treatments or trends over time)	The randomisation scheme prevented knowledge of upcoming assignments (clarified via email).	
Blinding of participants and personnel (3)	High risk of bias (Performance bias)	Blinding not mentioned in manuscript but email correspondence confirms patients and personnel were aware of allocations.	
Blinding of outcome assessment (4)	High risk of bias (Performance bias)	Blinding not mentioned in manuscript but email correspondence confirms outcome assessors were aware of allocations.	
Incomplete outcome data (5)	Low risk of bias (Attrition bias)	n=2 patients dropped out in first week, stating they were unable to complete the requirements of the protocol because of time constraints. Email correspondence confirms that they provided little data besides baseline data and were totally excluded from analyses.	When. subjects withdraw during the first condition of within-subjects design and provide no data for the second it is common to totally exclude them from analyses, as done here (Higgins et al., 2017). This may bias results if reasons relates to prognostic factors or outcomes (Mills et al., 2009). In this case reasons relate to general demands of the study (e.g. completing food diaries and blood draws) rather than aspects specific to one treatment arm. Attrition was also low.

Design feature (and number for guidance notes)	Risk of bias, confounding, imprecision or poor external validity. (unclear, low, high or n/a)	Explanation of decision, with explicit reference to methods used and measures taken to minimise impact.	Where risk is high or unclear, how important is this source of bias/ confounding and what is the likely direction and magnitude of its impact.
Selective outcome reporting (6)	Low risk of bias (selective reporting)	No evidence of selective reporting was noted after comparing the methods section to the results section and comparing the study to the registered protocol Some non-relevant outcomes described in the protocol are published in a separate study.	**NB** this column was largely redundant as the question was addressed in table 2
Carryover in within-subjects designs .e.g. crossovers (7)	Low risk of bias (carryover)	**Washout duration and protocol:** Extensive washout period: 3 weeks on usual formula. **Statistical tests for carryover:** ANOVA found no evidence of carryover or sequence effects. **Reason to expect carryover:** no logical reason to expect carryover as PKU is a stable, chronic condition that cannot spontaneously improve or deteriorate and GMP and formula have short-lived effects.	
Influence of funding source (8)	Low risk of bias	Study was funded via government grant but principal author has commercial links to GMP products at market and a manufacturer provided the GMP used. However, Cochrane and AHRQ guidance requires compelling evidence that funding sources or industrial links have increased the risk of specific sources of bias. There is no such evidence herein.	
Validity and reliability of outcome assessment (9)	Low risk (of measurement error)	(a) Measurements for main effect (comparing average change in plasma phe from baseline to study, between conditions) used venous blood samples analysed via an Hitachi L-8900 amino acid analyser with post–column ninhydrin derivatization. (b) Measurements for secondary analyses (comparing control at intermediate timepoints over each 3 week period) was via duplicate dried blood spots analysed via tandem MS (nonderivatized flow-injection method used for newborn screening). (c) all samples were collected from 9.00 – 9.30 after an overnight fast.	(a) Interassay CV for Hitachi L-8900 = 1.6% and systematic errors with modern methods such as this are comparable and inconsequential van Wegberg et al., 2017). (b) Phe content of bloodspots is generally 8-28% lower than plasma/ blood but this varies individually (van Wegberg et al., 2017) and the interassay CV using DBS and tandem MS is wide (~11.7%). However, the authors only compare blood spot data to itself and use it as secondary measure for triangulation.

Design feature (and number for guidance notes)	**Risk of bias, confounding, imprecision or poor external validity. (unclear, low, high or n/a)**	**Explanation of decision, with explicit reference to methods used and measures taken to minimise impact.**	**Where risk is high or unclear, how important is this source of bias/ confounding and what is the likely direction and magnitude of its impact.**
External validity (10)	High risk (of poor external validity)	**Completeness of GMP:** (a) A complete, commercial form of GMP was used, (Glytactin) recreating clinical practice. **Fairness of comparison:** (b) Modern and conventional versions of GMP were compared to modern and conventional versions of formula, a fair comparison that recreates clinical practice. **Setting:** (c) Subjects were outpatients, recreating clinical practice. **Sample selection and other threats to external validity:** Subjects sampled GMP before randomisation and those finding it unacceptable were excluded. Participants then self-selected GMP foods. This is (arguably) unlikely to occur in practice.	**Importance:** high **Direction and magnitude of influence:** Selectively recruiting patients that favour GMP would be expected to overstate GMPs benefit when considering the entire PKU population, some of whom may not prefer it.
Appropriateness of sample size and statistical analyses (11)	Low risk (of poor power/ precision)	**Power calculation:** (a) Power calculation was performed to ensure study was adequately powered: "The trial was powered at 80% (b = 0.20) based on a previously reported SD of 150 mmol/L for plasma Phe to detect a change in plasma phe of 120 mmol/L at a = 0.05. This provided a sample size of 25 subjects. 30 subjects were recruited with an estimated dropout rate of 15-20%." **Sample size:** see above **Appropriate analyses?** Appropriate analyses were used (ANOVA).	
Other study/ design-specific threats to quality, bias or validity (12)	Low risk	None	

6.2: Appendix 3: PRISMA-P Compliant protocol

Headings correspond to PRISMA-P requirements.

Methods that were changed after producing the protocol are ~~crossed out~~. Explanations are provided in ***bold italics.***

Item 1: Administrative information.

1a: Identification

The effect of replacing conventional amino-acid formulas with glycomacropeptide-based alternatives on phenylalanine control in adults and children with phenylketonuria: Protocol for a systematic review.

1b: Update

The review is original.

Item 2: Registration

A protocol adhering to PRISMA-P guidelines was produced before starting the review. This was undertaken to ensure transparency and prevent bias by ensuring methods and analyses were determined in advance, to plan the review process, to educate the first-time reviewer regarding good practice and to ensure items required for PRISMA reporting standards were addressed. The protocol was included in appendices and modifications/ deviations from it were noted. However, it was considered beyond the scope of the assignment to publish the protocol prospectively.

Item 3: Authors

3a: Contact information & 3b: Contributions

Contact 1: Mr Roderick Thomson, University of Chester, 1622331@chester.ac.uk, 8 Apsley Avenue Wallasey Merseyside

Role: MSc student, protocol author and sole reviewer.

Contact 2: Sohail Mushtaq, University of Chester

Role: Masters Dissertation supervisor.

Item 4: Amendments

Amendments to the protocol occurring after initiation of the review are inevitable as the nature of the evidence reviewed cannot always be anticipated. To ensure transparency, methods that were altered are included ~~but crossed out~~. Explanations are provided in ***bold italics.***

Item 5: Support

No financial support or sponsorship was received. Academic advice was received from the dissertation supervisor and the review was conducted by the sole author.

Item 6: Rationale

The in-born error of metabolism Phenylketonuria was associated with severe neurocognitive disabilities that usually required lifelong residential care before dietary treatment was widely adopted. Dietary treatment restricts dietary phenylalanine intake to 'control' plasma phenylalanine levels, preventing phe accumulating to toxic levels. As the developing CNS is most vulnerable to phe, treatment is initiated at birth and strict control (via strict dietary restriction) is essential during childhood to avoid the most severe defects. Some relaxation of control is permitted in adulthood but lifelong treatment is universally recommended as subtler neurocognitive defects occur if treatment ceases. Since the widespread adoption of dietary treatment, the severe disabilities that were previously associated with PKU have become rare. However, intelligence levels and several aspects of health remain suboptimal even among early-treated PKU patients, partly due to limitations of conventional dietary treatment. Adherence proves difficult as phe is ubiquitous in proteins so, in practice, most protein-containing foods must be restricted. Particularly

problematic are the synthetic amino-acid based protein substitutes (formulas) prescribed to meet outstanding protein requirements, which lack acceptability and satiety.

GMP is whole protein found in whey that is naturally free of phenylalanine. It is used to manufacture therapeutic foods that form a novel alternative to formulas and may improve dietary treatment and patients health in several ways. Evidence linking GMP to improvements in long-term health outcomes is lacking, owing to its recent development, the low prevalence of PKU and ethical concerns which make conducting trials difficult. Nevertheless, preliminary evidence suggests GMP is widely considered more acceptable than formula and that it may (for example) improve bone health, satiety and gastrointestinal comfort. However, GMP contains some phenylalanine as its purification from other phe-containing constituents of whey is imperfect. This residual phe raises concerns around the impact of GMP foods on control and their safety, particularly in children who are more sensitive to phe. Despite GMP being licenced in several countries for use in over 4s, these concerns remain unresolved. The scoping review conducted whilst planning this review attempted to resolve these concerns using narrative methods. Its findings suggested both adults and children can maintain control after replacing formula with GMP but that individualised titrations (adjusting the amount of formula replaced by GMP rather than simply replacing all formula) may be necessary in some children. Equivalent control would be a positive finding given GMPs other potential advantages. However, the review located only seven primary studies, most lacking randomisation and employing low quality, bias-prone, and heterogeneous designs. Other GMP reviews located used narrative methods and were concentrated within a single group with commercial links to GMP products, raising objectivity concerns. GMPs effect upon control in both adults and

children thus requires clarification using systematic methods that attempt to synthesize the entire evidence base and consider the impact of design quality, risk of bias and heterogeneity upon results.

Item 7: Objectives

This systematic review aims to clarify the effect of totally or partially replacing amino-acid based protein substitutes with GMP-based alternatives upon blood phenylalanine levels (I.E. 'control') in adults and children with phenylketonuria. As phe tolerance, protein requirements, formula requirements and several other aspects of care differ markedly between adults and children, the review separately analyses and presents data from studies involving adults (≥16) and children (3- 16). The following objectives are considered separately in both age groups.

Primary Objective: Through a systematic review using predetermined, evidence-based methods, determine the effect of replacing formula with GMP-based foods upon plasma phe control by narratively synthesizing studies and considering the impact of design quality, bias and heterogeneity upon their results.

Secondary Objective 1: Determine the influence of GMP upon compliance with therapeutic diets, and whether GMP-induced compliance changes lead to improvements, maintenance or deterioration of control. This will be accomplished by comparing compliance levels and control data between studies whilst considering the validity and reliability of methods used to monitor compliance.

~~**Secondary Objective 2:** Most studies compare newly developed ready-to-eat, food-like GMP products (e.g. snack bars) to patients usual formula. In studies where this formula is predominantly in traditional powdered formats, the comparison may be considered~~

misleading as more modern versions of formula are available. Clarify the impact of this upon findings using narrative sub-analyses.

Secondary Objective 2: Data permitting, clarify the effect upon plasma phe of only partially replacing formula with GMP. Among children, 100% replacement may be impossible due to low tolerances and high formula requirements. In adults the strategy may also prove useful (for example) in patients struggling to comply.

Secondary Objective 3: Data permitting, describe the extent of inter-individual variation in plasma phe responses or the amount of formula replaceable with GMP, and explore this variation as this may inform clinical practice by helping to identify patients likely to respond well to GMP.

The above objectives were addressed within the review but not as standalone objectives, as it was felt that so many objectives fragmented analyses particularly when the factors considered are all inter-related.

Secondary Objective 4 (children only): Consider whether GMP treatment remains worthwhile where only small quantities are permissible

Secondary Objective: 5 (children only) Consider whether improved purification techniques could reduce GMPs phe content, as this limits its utility most.

These objectives were removed as the data collected does not address them, and answering them essentially requires 'another review' something the handbook warns against

Item 8: Eligibility criteria

Rationale: Inclusion criteria were developed using PICOS to favour inclusivity and adherence of studies to them was ensured using screening and inclusion criteria forms. Rather than exclude studies during screening and eligibility assessments, the criteria selected considered most controlled human studies comparing GMP to formula acceptable. This involved incorporating more bias-prone and low-quality designs, though the quality/ risk of assessment process and subsequent narrative analyses considered the impact of design quality and bias upon findings. The approach also increased heterogeneity between studies, making it more difficulty to attribute changes in control to study-level or individual-level factors, as sources of heterogenity confound sub-analyses during narrative analysis. The approach was considered necessary despite these drawbacks given the few, low-quality, heterogeneous studies located in the scoping review.

Criteria: See screening and inclusion criteria forms for the criteria used.

Item 9: information sources

Guidance consulted: Information sources were selected and the search strategy formulated using Guidance from the Cochrane collaboration's handbook, the Centre for Reviews and Dissemination (CRD, 2009), the Agency for Healthcare Research and Quality (Relevo & Balshem, 2011) and the university information specialist.

The review aims to minimise publication bias whilst increasing sensitivity by searching a wide variety of sources of published and unpublished (or grey) literature.

Language restrictions: Only English language papers are include despite the possibility this may introduce language bias, as translation facilities are unavailable.

Date restrictions: no date restrictions are specified as records can be mis-indexed.

The resources searched where divided into database-type resources (searched first) and snowball-type resources that were identified from the results of the data-base type resources, after removal of duplicates and screening.

Database-type resources

See tables 12 -14 (appendix 1)

Snowball-type resources

As several of these resources were identified from the screened studies obtained from database-type resources, they could not be specified in advance.

1. **Reference list searches:** Included screened studies from database-type resource searches and reviews and guidelines located during the scoping review and from database-type resources
2. **Citation search tools:** titles of screened studies from database-type resource searches and reviews and guidelines located during the scoping review and from database-type resources were searched for
    a. Web of Science
    b. Google Scholar
    c. ~~Scopus~~ ***could not be used as host institution does not subscribe to service***
3. **Handsearching key journals:** *(all identified through database-type resources)*
    a. *Molecular genetics metabolism*
    b. *Journal of inherited metabolic disease*
    c. *American journal of clinical nutrition*

4. **Manual searches key websites** *(some were identified iteratively from other organisations sites)*
    a. Sites of companies producing GMP medical products sites (Biomarin, Cambrooke, Vitaflo UK, Vitaflo USA, Nutricia UK, Mead Johnson & Abbot Nutrition)
    b. PKU/ IEM society websites (e.g. ESPKU, NSPKU, NPKUA, Canadian PKU Association, PKU.IRE, New South Wales PKU Association, IEM Societies
    c. Dietetic. Nutrition Society Web sites (BDA, USA Academy of Nutrition & Dietetics, dietitians of Canada, Dietitians Association of Australia, Confederation of Dietetic Associations, Nutrition Society)
5. Ensuring all relevant studies have been identified by sending a list to experts
    a. Sent list to Professor Denise Ney

Item 10: Example Search Strategy

Rationale: The search strategy favours sensitivity over precision for aforementioned reasons. As varied non-randomised designs are included, search terms do not restrict results by design and relate only to the 'Condition' and 'Intervention' aspects of PICOS (Reeves et al., 2011). As per evidence-based guidelines (CRD, 2009 page 245) few concepts (condition and intervention) are combined using the Boolean operator 'AND' but many synonyms for each concept are grouped using 'OR' and truncation via wildcards. To increase sensitivity terms use natural language alongside database-specific subject headings (e.g. MeSH terms) which can be assigned inconsistently or slowly by indexers (Relevo & Balshem, 2011). Subject headings were 'exploded' where possible to include sub-topics in results.

A strategy specific to Pubmed was first devised by adapting the Cochrane Collaboration's in-born error of metabolism search strategy. This was pilot tested by ensuring it retrieved relevant articles known to be indexed therein. The syntax and terms of the strategy were then adapted for other databases and resources. The strategy and adaptation process were then checked against a peer-reviewed checklist (McGowan et al., 2015) and finalised through consultation with an information specialist.

Pubmed Search strategy: Phenylketonurias OR Phenylketonuria OR PKU OR "Phenylalanine hydroxylase deficiency" OR "PAH deficiency" OR hyperphenylalaninemia **AND** GMP OR Caseinomacropeptide OR Caseinomacropeptides OR Glycomacropeptide OR Glycomacropeptides OR "Protein substitute" OR "Protein substitutes" OR "amino acid formula" OR "phenylalnine free"

Adaptation Process for other resources

1. Check the database/ resource basic rules, controlled vocabulary & explosion rules
2. Check use of wildcards for truncation
3. Check rules for plurals, quotations and adjacency
4. Check 'tags' to increase precision if needed. E.g. [TIAB] & [TW] – some databases require them (WoS)
5. Adapt the strategy using 1-4, checking for reports of syntax errors etc.
6. Pilot the adapted strategy by searching for records known to be indexed therein

Item 11a: data management

Reference management software (Zotero) was used to maintain a database of studies throughout the entire review. The last date and details of searches were recorded in a table

that will be published. Database-type resources were searched first and results pooled in the software. Duplicate reports were removed automatically then manually and remaining reports were then screened using the screening form.

These screened reports were used to direct searches of snowball-type resources. Their reference lists were searched for unique reports and their titles were inserted into citation search tools, along with relevant reviews and guidelines located via database resources and during the preceeding review. Handsearches of journals were restricted to those that generated the most relevant reports during database searches. Potentially relevant reports from snowball-type resources were screened as they were encountered.

A final pool of screened reports was formed to detect different reports of the same study by comparing several study characteristics including authors, sample sizes and outcomes (CRD, 2009 page 25). Reports describing the same study were merged but all reports were retained to provide additional detail and check for publication bias. Where full text was unavailable (e.g. for proceedings or protocols) or clarification was required, authors were contacted via email which is quicker and may increase response rates (Relevo & Balshem, 2011). A window of seven days was allowed for responses.

Zotero was replaced by EndNote during searches of database-type resources, to improve compatibility with the bulk data export formats used by different databases.

Item 11b: Selection process

Abstracts and titles of articles were initially screened for inclusion using a screening form favouring inclusivity and speed of use. The full text of potentially relevant articles was then checked against a more detailed inclusion criteria form, though given the open nature of the review this also heavily favoured inclusivity. Both forms were developed a-priori using PICOS

to minimise bias, piloted using eligible, ineligible and borderline studies and are included in appendices. Whilst repetition of screening and study selection by a second, blinded reviewer with another speciality can minimise errors neither was possible.

A PRISMA flowchart is included in the review.

Item11c: Data collection process

A standard data extraction form was developed using evidence-based guidance a-priori to minimise bias then piloted using studies meeting the inclusion critera. It is included in appendicies. The form was developed and used alongside the quality assessment tool to ensure data required for quality assessment was also collected. A single reviewer extracted data though the process was double-checked to reduce errors. Where important details were missing, clarification was requested from authors via a standard email. If no response was received within seven days the information in question was marked 'unclear'. Studies providing insufficient information were excluded if this was subjectively deemed necessary, and this decision explained and its impact on findings considered.

Authors were allowed 18 days to respond to increase the likelihood of responses, not seven

Item 12: Data items collected.

See the data extraction form and quality assessment tool in appendices, where all data items are collected for are incorporated. Forms were developed in advance to avoid selective reporting and data items were selected using PICOS. As NRS are included, potential sources of confounding were identified. Where these related to participant-level characteristics in parallel studies (e.g. tolerance or history of control) their baseline

distribution between groups was investigated as imbalances may introduce selection bias. Sources of confounding relating to behaviours (e.g. dietary confounding or non-compliance) were assessed in separate domains of the quality assessment tool.

Item 13: Outcomes data is collected for

Primary outcomes: the mean difference in plasma phe (in micro moles/L) when comparing GMP treatment to control groups (usual treatment with AA formula), noting the amount of formula replaced by GMP (where applicable). The direction, magnitude, power, precision and clinical significance of effects will be noted.

Analysis of harm: is encompassed in the primary objective as decreases in control of sufficient magnitude are dangerous, particularly in childhood.

Secondary outcomes: Compliance levels with protein substitutes (quantity, frequency) and dietary requirements (protein and energy intakes)

Justification of using a surrogate: As neurocognitive defects such as impairments in intelligence and executive functions form the major source of morbidity in PKU, neurocognitive outcomes are most important. However, detectable changes in such outcomes accumulate over timespans that are impractically long to investigate. Historical prospective studies investigating the effect of dietary treatment upon such outcomes were plagued by poor compliance and patients 'crossing over' to conditions other than those they were initially randomised to (Poustie & Wildgoose, 2010). Given GMPs recent development no studies administering GMP located in the scoping review monitored such outcomes. Further, neurocognitive outcomes are assessed using a variety of indices that are difficult to compare (van Spronsen, Huijbregts, Bosch & Leuzzi, 2011). Whilst the pathophysiology of PKU is not totally understood, plasma phe concentration is the only physiological variable

implicated in every theoretical pathophysiological mechanism that is routinely measurable in research/ clinical settings and large meta-analyses show longer-term plasma phe levels robustly predicting neurocognitive outcomes (Waisbren et al., 2007; Fonnesbeck et al., 2012; Albrecht et al., 2009). Plasma phe concentration is thus most well-established and widely used marker in the study and treatment of PKU. Elevations are used to diagnose PKU in newborns and treatment guidelines advise the maintenance of levels associated with neurocognitive outcomes most resembling population norms through dietary phe restriction and blood monitoring (van Wegberg et al., 2017). Any validated assessment method (for blood/plasma phe) was considered acceptable but the impact of the validity and reliability of the method selected was evaluated during quality assessment and the narrative analysis.

Item 14: Assessing design quality and risk of bias

The risk of bias of included studies was assessed via quality assessment tool that was initially based on the Cochrane risk of bias tool. This assesses studies across six 'domains', each of which addresses a design feature which theory or empirical evidence suggests can introduce bias. (1) Random sequence generation (2) Concealment of allocation (3) Blinding of patients & personnel (4) Blinding of outcome assessments (5) Incomplete outcome data (6) Selective outcome reporting. For each domain, studies were assigned 'high', 'low' or 'unclear' risk of bias, using Cochrane guidance. Ratings were assigned only after requesting missing data from authors, as above. The tool was favoured for several reasons. It distinguishes conduct (which may introduce bias) from reporting (which may reflect details being omitted due to word counts). It also forces the reviewer to explain decisions and the weighting attached to each domain for each study in words, with explicit reference to the study manuscript. This ensures applicability, transparency and accountability.

Within-subjects designs (e.g. crossovers) are common in PKU to control for individual differences in baseline control and tolerance and to reduce sample size requirements, given the condition's low prevalence. The tool was adapted to evaluate crossover studies using Cochrane guidance (Higgins et al., 2011). An additional domain evaluating the risk of 'carryover', a form of bias unique to crossovers, was added and the domain regarding attrition was modified.

Further adaptations were necessary to evaluate NRS. Cochrane guidance directed these (Reeves et al., 2011). ~~As randomisation is not performed, parallel NRS are particularly susceptible to selection bias I.E. systematic differences in the distribution of potentially confounding factors between groups at baseline. Therefore, potential confounders (e.g. tolerance, age) were identified a-priori, and the data-extraction and quality assessment tool considered their distribution between groups at baseline and any attempts to compensate or balance groups.~~

Studies with inadequate randomisation were considered innately at risk of selection bias as statistical adjustments and efforts to balance prognostic factors between groups do not reliably prevent selection bias (Deeks et al., 2003), so the above was not performed.

In order to assess every aspect of study design quality at once the tool also evaluates design features that influence the confidence that can placed in a study's findings and their applicability, but that are not included in the Cochrane risk of bias tool as they do not relate to bias perse. These include factors affecting the power & precision of studies (which a meta-analysis would incorporate but a narrative analysis may overlook), factors affecting external validity and the extent to which compliance with treatment and dietary confounding affect findings. The Downs & Black checklist assisted when identifying these

factors (Downs & Black, 1998). A 'catchall' domain relating to design-specific sources of bias (e.g. recall bias in retrospective designs) was also included. Guidance notes were produced for the quality assessment tool. Where insufficient information to form a judgement was available for certain domains, clarification was requested from authors via email. If no response was received within 18 days, the domain was marked as unclear.

Item 15: Data synthesis

Given the bias-prone, heterogeneous studies located in the scoping review and that the direction of effect varied between studies, quantitatively pooling results would be misleading. It may obscure 'real' differences between results (due to heterogeneity) and may add unwarranted credibility to findings, given the bias-prone designs included. Instead, narrative analysis was performed using a structured approach based upon guidance from the CRD (2008, page 48) and particularly Popay et al., (2006). It was not possible to select the precise analysis methods to be used in advance. However, the variables analyses are to focus on were planned and listed below a-priori. Any analyses added afterwards as they were not anticipated or emerged from the data are marked separately and explained. This approach aims to avoid unnecessary post-hoc analyses or 'data-dredging' which in a quantities study would increase the chance of cumulative errors. In this narrative analysis they increase the likelihood of factors without plausible ways of influencing control being spuriously identified as important.

Analyses planned a-priori

(a) Separate analyses by age: As tolerance, the amount of formula replaceable by GMP and several other aspects of treatment differ markedly between adults and children, studies will

be separated to form separate reviews for adults for children. All subsequent analyses are performed in both groups separately.

NB: Note on Influence of heterogeneity on differences in effects between studies: Given the varied designs included, many sources of heterogeneity are likely to remain between studies despite dividing them by age. Examples include

(a) Samples with different severities of PKU, phe tolerances (both related – and inextricably linked to intensity of treatment as restriction and formula requirements are greater in severe cases)
(b) Differences in baseline control between samples
(c) Differences in patients and history of control/ compliance with treatment
(d) Use of different types of GMP (e.g. different LNAA content)
(e) Use of different types of formula

These sources of heterogeneity are likely to complicate/ confound sub-analyses attempting to attribute differences in effect estimates between studies (or differences in control between subjects, within studies) to factors suspected to explain them. To illustrate; studies divided into subgroups differing by one variable of interest (e.g. randomisation) to investigate its impact upon control are also likely to differ in the distribution of other variables that influence control (e.g. severity) confounding the comparison.

(b): Assessing the impact of bias and design quality upon the direction and magnitude of findings (primary objective 1): These comparisons should ascertain the impact of design quality/ bias upon findings which heavily influences how much confidence can be placed in the findings reported. As most forms of bias/ design limitations tend to overstate an

interventions benefits, more supportive (but less credible) findings would generally be expected from lower quality, bias-prone studies and less supportive (but more credible) findings from stronger studies. However, the likely impact of each form of bias/ design limitation will be considered individually.

Secondary objective 1: Investigating GMPs effect on compliance. One 'mechanism' through which GMP may improve control is through its widely-reported greater acceptability leading to improvements in compliance with the PKU diet in free-living subjects. This remains speculative however, and some studies suggesting GMP is preferable compare single servings of GMP and formula rather than complete formula-based and GMP-based diets, the latter of which contain fewer exchanges. A synthesis of study findings regarding compliance levels with GMP diets (and any resulting impacts upon control) that considers the validity and reliability of methods used to monitor compliance may clarify matters.

Assessing the strength of the evidence body reviewed for each age group, The strength of the entire evidence base (and thus the confidence that can be placed in the reviews findings) will be rated using GRADE (see final section).

Item 16: Meta-bias

The extent and likely impact of publication bias will be considered qualitatively, as funnel plots and statistical methods are inappropriate where heterogenity is pronounced, difficult to interpret and detect small study effects rather than publication bias perse.

Item 17: Confidence in cumulative estimate.

The quality of the entire body of evidence reviewed will be assessed using the GRADE system though adaptation was performed to permit its use in a narrative review.

6.0: References

Abdel-Salam, A. M., & Effat, L. K. (2010). Preparation and evaluation of a novel therapeutic dairy-based drink for phenylketonuria. *North American Journal of Medical Sciences*, 2(2), 66–70.

Ahring, K. K., Lund, A. M., Jensen, E., Jensen, T. G., Brøndum-Nielsen, K., Pedersen, M., … Møller, L. B. (2018). Comparison of glycomacropeptide with phenylalanine-free synthetic amino acids in test meals to PKU Patients: no significant differences in biomarkers, including plasma phe Levels. *Journal of Nutrition and Metabolism*, *2018*, 1–11.

AJCN (2017). American Journal of Clinical Nutrition: Instructions to authors. Retrieved from: https://academic.oup.com/ajcn/pages/General_Instructions

Albrecht, J., Garbade, S. F., & Burgard, P. (2009). Neuropsychological speed tests and blood phenylalanine levels in patients with phenylketonuria: A meta-analysis. *Neuroscience & Biobehavioral Reviews*, *33*(3), 414–421.

Blau, N., Hennermann, J. B., Langenbeck, U., & Lichter-Konecki, U. (2011). Diagnosis, classification, and genetics of phenylketonuria and tetrahydrobiopterin (BH4) deficiencies. Molecular Genetics and Metabolism, 104, S2–S9.

Blau, N., & Longo, N. (2015). Alternative therapies to address the unmet medical needs of patients with phenylketonuria. Expert Opinion on Pharmacotherapy, 16(6), 791–800.

Blau, N., van Spronsen, F. J., & Levy, H. L. (2010). Phenylketonuria. *The Lancet*, *376*, 1417-1427

Burgard, P., Bremer, H. J., Bührdel, P., Clemens, P. C., Mönch, E., Przyrembel, H., … Ullrich, K. (1999). Rationale for the German recommendations for phenylalanine level control

in phenylketonuria 1997. European Journal of Pediatrics, 158(1), 46–54.

Cambrooke (2018). Cambroke therapeutics UK: Glytactin BetterMilk information page and datacard. Retrieved from: http://cambrooketherapeutics.co.uk/glytactin-bettermilk-15/

Clarivate (2018). InCites journal citation reports: nutrition & dietetics category. Clarivate Analytics. Retrieved from: http://jcr.incites.thomsonreuters.com/JCRJournalHomeAction.action?pg=JRNLHOME&year=2017&edition=SCIE&categories=SA&newApplicationFlag=Y

Cochrane (2018). Cochrane collaboration inborn errors of metabolism trials register database subset search strategy: PKU subset. Retrieved from: https://cfgd.cochrane.org/inborn-errors-metabolism-trials-register

Crowe, M., & Sheppard, L. (2011). A review of critical appraisal tools show they lack rigor: Alternative tool structure is proposed. *Journal of Clinical Epidemiology*, *64*(1), 79–89.

CRD (2009). Systematic Reviews: CRD's guidance for undertaking reviews in healthcare. Centre for Reviews & Dissemination, University of York. Retrieved from https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf

Daly, A., Chahal, S., Evans, S., & MacDonald, A. (2016) *Impact of the amino acid profile of casein glycomacropeptide on metabolic control in children with PKU.* Abstract number P-099 presented at Society for the Study of Inborn Errors of Metabolism 2016 Annual Symposium: Rome, Italy, September 2016. *Journal of Inherited Metabolic Disease*, 39(S1), S85.

Daly, A., Evans, S., Chahal, S., Santra, S., & MacDonald, A. (2017). Glycomacropeptide in children with phenylketonuria: does its phenylalanine content affect blood phenylalanine control? *Journal of Human Nutrition and Dietetics*, *30*(4), 515–523.

Deeks, J. J., Higgins, J. P. T., & Altman D. G. (2017) Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins, J. P. T., Churchill, R., Chandler, J., & Cumpston, M. S. (Eds.), Cochrane Handbook for Systematic Reviews of Interventions Version 5.2.0 (Updated June 2017). *Cochrane.* Retrieved from https://training.cochrane.org/handbook/pdf-versions

Deeks, J., Dinnes, J., D'Amico, R., Sowden, A., Sakarovitch, C., Song, F., … Altman, D. (2003). Evaluating non-randomised intervention studies. *Health Technology Assessments*, *7*(27), 1-173.

DeRoche, K., & Welsh, M. (2008). Twenty-Five Years of Research on Neurocognitive Outcomes in Early-Treated Phenylketonuria: Intelligence and Executive Function. *Developmental Neuropsychology*, *33*(4), 474–504.

Downs, S. H., & 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 & Community Health, 52(6), 377–384.

Dwan, K., Gamble, C., Williamson, P. R., Kirkham, J. J., & the Reporting Bias Group. (2013). Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias — An Updated Review. *PLoS ONE*, *3*(8), e3081.

Egger, M., Zellweger-Zähner, T., Schneider, M., Junker, C., Lengeler, C., & Antes, G. (1997). Language bias in randomised controlled trials published in English and German. *Lancet*, 350(9074), 326–329.

Enns, G. M., Koch, R., Brumm, V., Blakely, E., Suter, R., & Jurecki, E. (2010). Suboptimal outcomes in patients with PKU treated early with diet alone: Revisiting the evidence. *Molecular Genetics and Metabolism*, *101*(2–3), 99–109.

Fonnesbeck, C. J., McPheeters, M. L., Krishnaswami, S., Lindegren, M. L., & Reimschisel, T. (2013). Estimating the probability of IQ impairment from blood phenylalanine for phenylketonuria patients: a hierarchical meta-analysis. *Journal of Inherited Metabolic Disease*, *36*(5), 757–766.

Giovannini, M., Verduci, E., Salvatici, E., Paci, S., & Riva, E. (2012). Phenylketonuria: nutritional advances and challenges. Nutrition & Metabolism, 9(1), 7.

Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., … deBeer, H. (2011). GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology*, *64*(4), 383–394.

Harding, C. O., & Blau, N. (2010). Advances and challenges in phenylketonuria. Journal of Inherited Metabolic Disease, 33(6), 645–648.

Ho, G., & Christodoulou, J. (2014). Phenylketonuria: translating research into novel therapies. Translational Pediatrics, 3(2), 14.

Higgins, J. P. T., Altman D. G., & Sterne, J. A. C. (2017) Chapter 8: Assessing risk of bias in included studies. In: Higgins, J. P. T., Churchill, R., Chandler, J., & Cumpston, M. S. (Eds.), Cochrane Handbook for Systematic Reviews of Interventions Version 5.2.0 (Updated June 2017). *Cochrane*. Retrieved from https://training.cochrane.org/handbook/pdf-versions

Higgins, J. P. T., Churchill, R., Chandler, J., & Cumpston, M. S. (2017) Cochrane Handbook for Systematic Reviews of Interventions Version 5.2.0 (Updated June 2017). *Cochrane*. Retrieved from https://training.cochrane.org/handbook/pdf-versions

Higgins, J. P. T., & Deeks, J. J., (2011) Chapter 7: Selecting studies and collecting data. In: Higgins, J. P. T., & Green, S. (Eds.), Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (Updated March 2011). *Cochrane*. Retrieved from

https://handbook-5-1.cochrane.org/

Higgins, J. P. T., Deeks, J. J., & Altman, D. G. (2011) Chapter 16: Special topics in statistics. In: Higgins, J. P. T., & Green, S. (Eds.), Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (Updated March 2011). *Cochrane*. Retrieved from https://handbook-5-1.cochrane.org/

Koch, R., Burton, B., Hoganson, G., Peterson, R., Rhead, W., Rouse, B., … Azen, C. (2002). Phenylketonuria in adulthood: A collaborative study. *Journal of Inherited Metabolic Disease.*, 25, 333-346

LaClair, C. E., Ney, D. M., MacLeod, E. L., & Etzel, M. R. (2009). Purification and Use of Glycomacropeptide for Nutritional Management of Phenylketonuria. *Journal of Food Science*, *74*(4), E199–E206.

Lau, J., Ioannidis, J. P. A., Terrin, N., Schmid, C. H., & Olkin, I. (2006). The case of the misleading funnel plot. *British Medical Journal*, *333*(7568), 597–600.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., … Moher, D. (2009). The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *British Medical Journal*, *339,* b2700.

Lim, K., van Calcar, S. C., Nelson, K. L., Gleason, S. T., & Ney, D. M. (2007). Acceptable low-phenylalanine foods and beverages can be made with glycomacropeptide from cheese whey for individuals with PKU. *Molecular Genetics and Metabolism*, *92*(1–2), 176–178.

Macdiarmid, J., & Blundell, J. (1998). Assessing dietary intake: Who, what and why of under-reporting. *Nutrition Research Reviews*, *11*(02), 231.

MacDonald, A., Harris, G., Rylance, G., Asplin, D., & Booth, I. W. (1997). Abnormal feeding

behaviours in phenylketonuria. *Journal of Human Nutrition and Dietetics*, *10*(3), 163–170.

MacDonald, A., Lilburn, M., Cochrane, B., Davies, P., Daly, A., Asplin, D., … Lee, P. (2004). A new, low-volume protein substitute for teenagers and adults with phenylketonuria. *Journal of Inherited Metabolic Disease*, *27*(2), 127–135.

MacDonald, Anita. (2000). Diet and compliance in phenylketonuria. *European Journal of Pediatrics*, *159*(S2), S136–S141.

MacDonald, Anita, Gokmen-Ozel, H., van Rijn, M., & Burgard, P. (2010). The reality of dietary compliance in the management of phenylketonuria. *Journal of Inherited Metabolic Disease*, *33*(6), 665–670.

MacLeod, E. L., Clayton, M. K., van Calcar, S. C., & Ney, D. M. (2010). Breakfast with glycomacropeptide compared with amino acids suppresses plasma ghrelin levels in individuals with phenylketonuria. *Molecular Genetics and Metabolism*, *100*(4), 303–308.

MacLeod, E. L., & Ney, D. M. (2010). Nutritional Management of Phenylketonuria. *Annales Nestlé (English Ed.)*, *68*(2), 58–69.

Matalon, R., Michals-Matalon, K., Bhatia, G., Burlina, A. B., Burlina, A. P., Braga, C., … Guttler, F. (2007). Double blind placebo control trial of large neutral amino acids in treatment of PKU: Effect on blood phenylalanine. *Journal of Inherited Metabolic Disease*, *30*(2), 153–158.

Matalon, R., Michals-Matalon, K., Bhatia, G., Grechanina, E., Novikov, P., McDonald, J. D., … Guttler, F. (2006). Large neutral amino acids in the treatment of phenylketonuria (PKU). *Journal of Inherited Metabolic Disease*, *29*(6), 732–738.

McGowan, J., Sampson, M., Salzwedel, D. M., Cogo, E., Foerster, V., & Lefebvre, C. (2016).

PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *Journal of Clinical Epidemiology*, *75*, 40–46.

Mills, E. J., Chan, A.-W., Wu, P., Vail, A., Guyatt, G. H., & Altman, D. G. (2009). Design, analysis, and presentation of crossover trials. *Trials*, *10*(27).

Moher, D., Jadad, A. R., Nichol, G., Penman, M., Tugwell, P., & Walsh, S. (1995). Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists. *Controlled Clinical Trials*, *16*(1), 62–73.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine*, 6(7), 6.

Mönch, E., Herrmann, M. E., Brösicke, H., Schöffer, A., & Keller, M. (1996). Utilisation of amino acid mixtures in adolescents with phenylketonuria. *European Journal of Pediatrics*, *155*(S1), S115–S120.

Mulrow, C. D. (1994). Systematic Reviews: Rationale for systematic reviews. *BMJ*, *309*(6954), 597–599.

Ney, D. M., Gleason, S. T., van Calcar, S. C., MacLeod, E. L., Nelson, K. L., Etzel, M. R., … Wolff, J. A. (2009). Nutritional management of PKU with glycomacropeptide from cheese whey. *Journal of Inherited Metabolic Disease*, *32*(1), 32–39.

Ney, Denise M., Blank, R. D., & Hansen, K. E. (2013). Advances in the nutritional and pharmacological management of phenylketonuria: *Current Opinion in Clinical Nutrition and Metabolic Care*, 17(1), 61–68.

Ney, Denise M, Stroup, B. M., Clayton, M. K., Murali, S. G., Rice, G. M., Rohr, F., & Levy, H. L. (2016). Glycomacropeptide for nutritional management of phenylketonuria: a randomized, controlled, crossover trial. *The American Journal of Clinical Nutrition*,

104(2), 334–345.

Ney, Denise Marie, & Etzel, M. R. (2017). Designing medical foods for inherited metabolic disorders: why intact protein is superior to amino acids. *Current Opinion in Biotechnology*, *44*, 39–45.

Odgaard-Jensen, J., Vist, G. E., Timmer, A., Kunz, R., Akl, E. A., Schünemann, H., … Oxman, A. D. (2011). Randomisation to protect against selection bias in healthcare trials. *Cochrane Database of Systematic Reviews*. Issue 4, Art. No.: MR000012.

Oxman, A. D., & Guyatt, G. H. (1993). The Science of Reviewing Research. *Annals of the New York Academy of Sciences*, 703, 125–134.

Paine, R. S. (1957). The variability in manifestations of untreated patients with phenylketonuria (phenylpyruvic aciduria). *Pediatrics*, 20(2), 290–302.

Pinto, A., Almeida, M. F., Ramos, P. C., Rocha, S., Guimas, A., Ribeiro, R., … Rocha, J. C. (2017). Nutritional status in patients with phenylketonuria using glycomacropeptide as their major protein source. *European Journal of Clinical Nutrition*, *71*(10), 1230–1234.

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., ... & Duffy, S. (2006). Guidance on the conduct of narrative synthesis in systematic reviews. *ESRC,* Retrieved from: http://www.lancaster.ac.uk/shm/research/nssr/research/dissemination/publications/NS_Synthesis_Guidance_v1.pdf

Poustie, V. J., & Wildgoose, J. (2010). Dietary interventions for phenylketonuria. The Cochrane Database of Systematic Reviews, CD001304.

Prince, A. P., Mcmurry, M. P., & Buist, N. R. M. (1997). Treatment products and approaches for phenylketonuria: improved palatability and flexibility demonstrate safety,

efficacy and acceptance in US clinical trials. Journal of Inherited Metabolic Disease, 20, 486– 498

Reeves, B. C., Deeks, J. J., Higgins, J. P. T., & Wells, G. A. (2011) Chapter 13: Including non-randomised studies. In: Higgins, J. P. T., & Green, S. (Eds.), Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (Updated March 2011). *Cochrane*. Retrieved from https://handbook-5-1.cochrane.org/

Relevo, R., & Balshem, H. (2011). Finding evidence for comparing medical interventions: AHRQ and the Effective Health Care Program. *Journal of Clinical Epidemiology*, *64*(11), 1168–1177.

Rocha, J. C., & MacDonald, A. (2016). Dietary intervention in the management of phenylketonuria: current perspectives. Pediatric Health, Medicine and Therapeutics, Volume 7, 155–163.

Savović, J., Jones, H., Altman, D., Harris, R., Jűni, P., Pildal, J., … Sterne, J. (2012). Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. *Health Technology Assessment*, *16*(35).

Schulz, B., & Bremer, H. J. (1995). Nutrient intake and food consumption of adolescents and young adults with phenylketonuria. Acta Paediatrica, 84(7), 743–748.

Schünemann, H., Brożek, J., Guyatt, G., & Oxman, A. (2013). GRADE Handbook: Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Retrieved from https://gdt.gradepro.org/app/handbook/handbook.html#h.m9385o5z3li7

Schünemann H. J., Oxman A. D., Vist G. E., Higgins J. P. T., Deeks, J. J, Glasziou P., … Guyatt, G. H. (2017) Chapter 12: Interpreting results and drawing conclusions. In: Higgins, J.

P. T., Churchill, R., Chandler, J., & Cumpston, M. S. (Eds.), Cochrane Handbook for Systematic Reviews of Interventions Version 5.2.0 (Updated June 2017). *Cochrane*. Retrieved from https://training.cochrane.org/handbook/pdf-versions

Solverson, P., Murali, S. G., Litscher, S. J., Blank, R. D., & Ney, D. M. (2012). Low Bone Strength Is a Manifestation of Phenylketonuria in Mice and Is Attenuated by a Glycomacropeptide Diet. *PLoS ONE*, *7*(9), e45165.

Song, F., Eastwood, A. J., Gilbody, S., Duley, L., & Sutton, A. J. (2000). Publication and related biases. *Health Technology Assessments*, 4(10), 1–115.

Smith, I., Cockburn, F., Barwell, B., Brenton, D., Chapple, J., Clark, B., … Laing, S. (1993). Recommendations on the dietary management of phenylketonuria. Archives of Disease in Childhood, 68(3), 426–427.

Spécola, N., & Chiesa, A. (2017). Alternative Therapies for PKU. *Journal of Inborn Errors of Metabolism and Screening*, *5*, 1-5

Suarez-Almazor, M. E., Belseck, E., Homik, J., Dorgan, M., & Ramos-Remus, C. (2000). Identifying Clinical Trials in the Medical Literature with Electronic Databases: MEDLINE Alone Is Not Enough, *Controlled Clinical Trials, 21, 476-487.*

van Calcar, S. C., MacLeod, E. L., Gleason, S. T., Etzel, M. R., Clayton, M. K., Wolff, J. A., & Ney, D. M. (2009). Improved nutritional management of phenylketonuria by using a diet containing glycomacropeptide compared with amino acids. *The American Journal of Clinical Nutrition*, *89*(4), 1068–1077.

van Calcar, S. C., & Ney, D. M. (2012). Food Products Made with Glycomacropeptide, a Low-Phenylalanine Whey Protein, Provide a New Alternative to Amino Acid–Based Medical Foods for Nutrition Management of Phenylketonuria. *Journal of the Academy of Nutrition and Dietetics*, *112*(8), 1201–1210.

van Spronsen, F. J., & Enns, G. M. (2010). Future treatment strategies in phenylketonuria. *Molecular Genetics and Metabolism*, *99*, S90–S95.

van Spronsen, F. J., van Wegberg, A. M., Ahring, K., Bélanger-Quintana, A., Blau, N., Bosch, A. M., … MacDonald, A. (2017). Key European guidelines for the diagnosis and management of patients with phenylketonuria. The Lancet Diabetes & Endocrinology, 5(9), 743–756.

van Wegberg, A. M. J., MacDonald, A., Ahring, K., Bélanger-Quintana, A., Blau, N., Bosch, A. M., … van Spronsen, F. J. (2017). The complete European guidelines on phenylketonuria: diagnosis and treatment. *Orphanet Journal of Rare Diseases*, *12*, 162

Vockley, J., Andersson, H. C., Antshel, K. M., Braverman, N. E., Burton B. K., Frazier, D. M., … Berry, S. A. (2014). The American College of Medical Genetics and Genomics Therapeutic Committee. Phenylalanine hydroxylase deficiency: diagnosis and management guideline. Genetics in Medicine, 16(2), 188–200.

Vitaflo UK (2017). PKU Sphere GMP product information page and datacard. Retrieved from: https://www.nestlehealthscience.co.uk/vitaflo/conditions/protein%20metabolism/phenylketonuria/pku-sphere

Waisbren, S. E., Noel, K., Fahrbach, K., Cella, C., Frame, D., Dorenbaum, A., & Levy, H. (2007). Phenylalanine blood levels and clinical outcomes in phenylketonuria: A systematic literature review and meta-analysis. *Molecular Genetics and Metabolism*, *92*, 63–70.

Walter, J., White, F., Hall, S., MacDonald, A., Rylance, G., Boneh, A., … Vail, A. (2002). How practical are recommendations for dietary control in phenylketonuria? *The Lancet*, *360*(9326), 55–57.

WHO (2007). Protein and amino acid requirements in human nutrition: report of a joint

WHO/FAO/UNU Expert Consultation [Geneva, 9 - 16 April 2002]. Geneva: WHO.

Yi, S. H., & Singh, R. H. (2015). Protein substitute for children and adults with phenylketonuria. Cochrane Database of Systematic Reviews, CD004731.

Zaki, O. K., El-Wakeel, L., Ebeid, Y., Ez Elarab, H. S., Moustafa, A., Abdulazim, N., … Elghawaby, A. (2016). The Use of Glycomacropeptide in Dietary Management of Phenylketonuria. Journal of Nutrition and Metabolism, 2016, 1–5.