

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email editorial.bmjopen@bmj.com

BMJ Open

Study Protocol for Developing a Core Outcome Set for Lifestyle Weight Management Programmes by Expert Consensus

| | |
|-------------------------------|---|
| Journal: | <i>BMJ Open</i> |
| Manuscript ID | bmjopen-2018-025193 |
| Article Type: | Protocol |
| Date Submitted by the Author: | 03-Jul-2018 |
| Complete List of Authors: | Mackenzie, Ruth; University of Glasgow, Institute of Cardiovascular and Medical Sciences Ells, Louisa; Teeside University, School of Health and social Care Simpson, Sharon; University of Glasgow, MRC/CSO Social and Public Health Sciences Unit Logue, Jennifer; University of Glasgow, Institute of Cardiovascular and Medical Sciences; |
| Keywords: | core outcomes, lifestyle weight management, Delphi |
| | |

SCHOLARONE™
Manuscripts

1
2
3 **Study Protocol for Developing a Core Outcome Set for Lifestyle Weight Management Programmes**
4 **by Expert Consensus**
5

6 *Ruth M. Mackenzie¹, Louisa J. Ells², Sharon Anne Simpson³, Jennifer Logue¹*
7

8
9 ¹ Institute of Cardiovascular and Medical Sciences, University of Glasgow
10

11 ² School of Health & Social Care, Teeside University
12

13 ³ Institute of Health and Wellbeing, University of Glasgow
14
15

16
17
18 Corresponding Author:
19

20 Jennifer Logue
21

22
23 Institute of Cardiovascular and Medical Sciences
24

25 University of Glasgow
26

27
28 126 University Place
29

30 Glasgow
31

32 G12 8TA
33

34
35 Jennifer.Logue@glasgow.ac.uk
36
37
38
39

40 **ABSTRACT**
41

42 **Introduction:** Weight management interventions in research studies and in clinical practice differ in
43 length, advice, frequency of meetings, staff, and cost. Very few real-world programmes have
44 published outcomes for patients, and those that have published used different ways of reporting the
45 information, making it impossible to compare interventions and further develop the evidence base.
46 Developing a core outcome set for weight management interventions will allow different weight
47 management programmes to be compared and reveal which interventions work best for which
48 members of the population.
49

50
51 **Methods and analysis:** An expert group, comprised of 40 people who work in, refer to, or attend
52 weight management programmes, will be asked to decide which outcomes services should report.
53
54
55
56
57
58
59
60

1
2
3 An online Delphi process will be employed to help the group reach consensus as to which outcomes
4 should be measured and reported, and which definitions/instruments should be utilised in order to
5 do so. The first stage of the Delphi process (3 rounds of questionnaires) will focus on outcomes while
6 the second stage (3 additional rounds of questionnaires) will focus on definition/instrument
7 selection.
8
9

10
11 **Ethics and dissemination:** Ethical approval for this study has been received from the University of
12 Glasgow College of Medical, Veterinary and Life Sciences Ethics Committee. With regard to
13 disseminating results, a report will be submitted to our funding body, the Chief Scientist Office of the
14 Scottish Government Health Department. In addition, early findings will be shared with Public Health
15 England (PHE) and Health Scotland, and results communicated via conference presentations, peer
16 review publication and our institutions' social media platforms.
17
18

19
20 **Registration details:** The project has been registered with the COMET (Core Outcome Measures in
21 Effectiveness Trials) Initiative (<http://www.comet-initiative.org/studies/details/1056>).
22
23
24
25
26
27

28 **STRENGTHS AND LIMITATIONS OF THIS STUDY**

- 29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
- The major strength of this study is that it is the first of its kind and development of a core outcome set for lifestyle weight management programmes is much needed in order to standardise reporting which, in turn, will lead to a better evidence base and improvements in weight management provision.
 - It is a limitation that this study is wholly based in the United Kingdom (UK) as the results may need some minor adaptation to be suited to real-world programmes set within other healthcare systems.
 - However, we will use the internationally recognised Delphi method to garner opinions from a wide range of individuals with expertise in lifestyle weight management.

48 **INTRODUCTION**

49
50 Both the National Institute for Health and Care Excellence (NICE)¹ and Scottish Intercollegiate
51 Guidelines Network (SIGN)² guidelines outline the intervention components to be included in a
52 community weight management programme, namely calorie restriction, increased physical activity,
53 and behavioural interventions. These have proven efficacy from randomised controlled trials³.
54
55
56
57
58
59
60

1
2
3 However, their implementation in practice is inconsistent with mapping exercises in Scotland⁴ and
4 England⁵ showing wide variation in services in terms of inclusion criteria, referral routes, delivery
5 format, length and cost. Few real life services have published data and when they do publish, results
6 can be poor with low levels of completion and 'success', and lack of longer term outcomes.
7
8

9
10 The NICE guidance, 'Weight management: lifestyle services for overweight or obese adults'¹,
11 identified a number of evidence gaps. These included, reliance on studies with short follow-up,
12 collection of data at limited time points, small sample sizes, demographic samples that limit the
13 ability to generalise, non-reporting of reasons for people dropping out and lack of evidence
14 regarding the effect of population characteristics, such as age, gender and socio economic status, on
15 the effectiveness of a service. They noted a lack of comparisons between lifestyle weight
16 management programmes in the United Kingdom (UK). This lack of an evidence base means that it is
17 not possible to issue clear guidance as to which services are cost effective for which population
18 groups.
19
20
21
22
23

24
25 Public Health England (PHE) has created a standard evaluation framework (SEF)⁶ to aid the
26 evaluation of real world weight management programmes. However, in their 2015 weight
27 management mapping exercise⁵, PHE reported that only 46% of adult weight management
28 programmes use the SEF and, as it simply suggests areas for reporting and potential methods of
29 analysis, there is a huge gap in standardised reporting. PHE had intended to analyse data from
30 services but analysis was not possible due to the heterogeneity of reporting which included
31 kilograms, % weight loss, average number of completers achieving 5% weight loss, body mass index
32 (BMI) and more⁵. With regard to research studies, evidence suggests similar heterogeneity in terms
33 of the reporting of outcomes⁷.
34
35
36
37
38

39
40 In an attempt to address this reporting issue, PHE issued a minimum dataset⁸ which provides an
41 important core outcome recommendation for England, stipulating collection of certain
42 demographics, service details, BMI and wellbeing at baseline, on completion of the programme and
43 at 6 months and 12 months post programme. A data collection tool provides information to support
44 the standardisation of these data collection practices. This minimum dataset will be used to support
45 PHE's recently released document on adult tier 2 weight management service key performance
46 indicators (KPIs)⁹ which provides advice as to how weight status and service compliance should be
47 reported and measured.
48
49
50
51
52

53 The study described herein, Developing Core Outcome Measures for Lifestyle Weight Management
54 Programmes by Expert Consensus, has been funded through a Chief Scientist Office of the Scottish
55
56
57
58
59

1
2
3 Government Health Department grant and will serve to further validate and build upon the PHE
4 minimum dataset⁸ and KPI document⁹, while also informing a similar framework for Scotland. In
5 addition, our research will provide much needed consensus on the measurements that should be
6 used, such as questionnaires, something currently not covered in the PHE minimum dataset⁸ or KPI
7 document⁹. Overall, this work will ensure more consistency in the measurement of the effectiveness
8 of weight management services, leading to a better evidence base from which to identify which
9 services are effective across a range of settings.
10
11

12
13
14 Therefore, the aim of this study, which will run from November 2017 until November 2018, is to gain
15 expert consensus opinion on the core outcomes that should be reported from lifestyle weight
16 management interventions in real world clinical practice as well as within research studies.
17
18

19
20 The specific study objectives are to:

- 21
22 1. Review the list of outcomes previously reported in the PHE SEF⁶, minimum dataset⁸ and KPI
23 document⁹;
- 24
25 2. Identify additional outcomes reported in studies of structured, sustained, multi-component
26 weight management programmes from a systematic review of the literature;
- 27
28 3. Select outcomes for inclusion in the core dataset using consensus methodology;
- 29
30 4. Select instruments for measuring chosen outcomes using consensus methodology.
31
32
33
34
35
36
37

38 **METHODS AND ANALYSIS**

39 **Identification of outcomes**

40
41
42 We will generate a list of outcomes by review of the PHE SEF⁶, which was itself developed from a
43 systematic review of the literature/focus groups, and from the PHE minimum dataset⁸ and KPI
44 document⁹ which were developed through expert consensus and evidence from the peer review and
45 grey literature.
46
47
48

49
50 Further outcomes will be selected by a review of included studies in the systematic review, 'The
51 clinical effectiveness of long-term weight management schemes for adults' by Hartmann-Boyce *et al.*
52 (2013)⁷, conducted during the development of NICE guidance¹. This systematic review⁷ assessed the
53 effects of multicomponent behavioural weight management programmes (BWMPs) in overweight
54 and adults with obesity which may be applicable in the UK. To be considered a multicomponent
55
56
57
58
59
60

1
2
3 BWMP, the components of the programme had to include diet, physical activity and behavioural
4 therapy (for example, counselling sessions). The scope included commercial weight loss programmes
5 and non-commercial programmes, such as those delivered in primary care settings (for example, in
6 GP practices)⁷. It updated and expanded on an existing systematic review published in 2011 by
7 Loveman *et al.*³ and used similar methods. The Loveman systematic review³ sought to assess the
8 long-term clinical effectiveness and cost-effectiveness of multicomponent weight management
9 schemes for adults in terms of weight loss and maintenance of weight loss.
10
11
12
13

14 Additional outcomes will be identified by updating the Hartmann-Boyce systematic review⁷, using
15 the same inclusion criteria but extending search dates so that studies from 1/11/2012 until 30/09/17
16 are included. Search and selection criteria for the systematic review are identical to those of
17 Hartmann-Boyce⁷. With regard to database searches, Hartmann-Boyce⁷ searched BIOSIS, the
18 Cochrane Database of Systematic Reviews, CENTRAL, the Conference Proceedings Citation Index, the
19 Database of Abstracts of Reviews and Effects (DARE), Embase, the Health Technology Assessment
20 database, Medline, PsychInfo, and Science Citation Index for references relating to weight loss
21 programmes. They also screened references from three additional sources: reference lists in
22 systematic reviews, documents received via the NICE call for evidence, and studies excluded from
23 Loveman³ that they wished to re-examine. Studies selected for inclusion had to be structured
24 sustained multi-component weight management programmes with interventions which were a
25 combination of diet and physical activity with a behaviour change strategy to influence lifestyle. In
26 addition, programmes were required to include a follow-up of more than 12 months and be
27 delivered in the health sector, in the community or commercially (i.e. applicable to the NHS).
28
29
30
31
32
33
34
35
36

37 Two review authors will independently assess the abstracts of studies resulting from our literature
38 search. Full text copies of studies appearing to meet the inclusion criteria will be further
39 independently assessed by the 2 reviewers. Following discussion, agreement will be reached as to
40 which studies to include. Any new outcomes will then be identified from the selected studies from
41 both Hartmann-Boyce⁷ and the updated review.
42
43
44
45
46
47

48 **Identification of Instruments**

49
50
51 By review of the studies identified during the systematic reviews previously described, we will list
52 instruments and definitions for selected outcomes. The study investigators will review this list and
53 add any further suitable instruments.
54
55
56
57
58
59
60

Data Analysis and Presentation

For analysis purposes, the data will be tabulated so that the outcomes and instruments to be included in our Delphi are listed and the study/studies from which they were identified are displayed. Outcomes and instruments will be grouped under appropriate domains following review of selected outcomes.

Participants

We will develop our core outcome set by means of consensus from an expert group. The sampling frame will aim to include members of the public with experience of NHS, local authority or commercial weight management programmes in the U.K., academics/policy makers/commissioners working in weight management, staff currently involved in delivering a lifestyle weight management programme for adults (without significant policy involvement), and primary care staff (referrers). There is no published agreement on the optimal size of an expert group; pragmatism is required while ensuring a range of opinions is garnered. Experience suggests a greater than 80% completion rate of Delphi questionnaires^{10;11}.

We will pre-approach potential volunteers to get agreement to participate from 10 members of the public, 20 academics/policy makers/commissioners, 20 weight management staff and 10 primary care staff. Forty experts will complete each of the two separate Delphi processes. For the first Delphi process (stage 1, outcome selection), 10 members of the public, 10 academics/policy makers/commissioners, 10 weight management staff and 10 primary care staff will be invited to participate. For the second Delphi (stage 2, instrument selection), 20 academics/policy makers/commissioners and 20 weight management staff will be invited to participate with further members recruited if any of the original group (the 10 from each group who completed stage 1) have dropped out after the stage 1 Delphi.

A small monetary incentive (a £35 gift voucher for either John Lewis or Amazon, depending on preference) will be offered to members of the public and primary care staff as this study is not of any direct benefit to them and could not be considered part of their role.

1
2
3 Staff working in weight management, academics/policy makers/commissioners and primary care
4 staff will be recruited by email from the investigators and their personal contacts, and also via an
5 email from the Association for the Study of Obesity. An information letter outlining the study will be
6 attached to emails. On registering interest in our study, we will ask volunteers from these groups to
7 provide us with information as to their role and geographical location within the UK.
8
9

10
11 Members of the public will be recruited by email from the Association for the Study of Obesity
12 (which has lay members) and from professional contacts (a number of weight management
13 programmes have lay members on steering committees). An information letter outlining the study
14 will be attached to emails. (The information letter for the public will be written in lay language and
15 will therefore differ slightly to the information letter for the other groups.) We have also registered
16 with the NIHR People in Research website (<https://www.peopleinresearch.org/>) where our study will
17 be advertised. Our information letter will be available to download from this website. On registering
18 interest in our study, a 'job description' pro forma will be sent to members of the public via email.
19 They will be asked to complete this pro forma and return it to us by email. The pro forma will
20 provide us with information as to their gender, age, geographical location and experience of weight
21 loss programmes.
22
23
24
25
26
27
28

29
30 In addition, Facebook and Twitter will be used to recruit members of the public, weight
31 management staff, academics/policy makers/commissioners and primary care staff. Facebook posts
32 and Tweets will link to a Mailchimp recruitment page where volunteers will be able to register their
33 interest. On doing so, they will receive the appropriate information letter. Weight management
34 staff, academics/policy makers/commissioners and primary care staff will be asked to provide us
35 with information as to their role and geographical location within the UK, and members of the public
36 will be asked to complete the job description pro forma.
37
38
39
40

41
42 Following provision of information regarding role and geographical location from weight
43 management staff, academics/policy makers/commissioners and primary care staff, and the return
44 of completed pro formas from members of the public, selection of volunteers to participate will
45 commence. Selection will be based on our sampling framework which is outlined below. Volunteers
46 will be sent an email to thank them for their interest and inform them if they have been selected to
47 participate or not. A list of selected volunteers' names and email addresses will then be sent to
48 Clinvivo (www.clinvivo.com, a spin-out company of the University of Warwick) who will be
49 conducting the Delphi process. Clinvivo will then contact these individuals by email, providing a link
50 to the online Delphi questionnaire and instructions as to how to complete it.
51
52
53
54
55
56
57
58
59
60

Sampling Framework

To ensure our volunteers are a representative UK group, of the 20 weight management staff selected, at least 50% will be from England. Similarly, at least 50% of the 20 academic/policy maker/commissioner group will be from England. 8 of the 20 (40%) will be academics, 6 of the 20 (30%) will be policy makers and 6 of the 20 (30%) will be commissioners. At least 50% of the 10 primary care staff selected will also be from England. With regard to members of the public, more than 50% will have experience of commercial weight loss programmes, more than 50% will be of working age, more than 30% will be male and less than 30% will be from any one region of the UK.

Delphi Survey

In order to develop our core outcome dataset, Delphi methodology will be used to gain consensus from our expert group. Two Delphis (stage 1 and stage 2) will be carried out using an online system developed and conducted by Clinvivo. Each Delphi will be carried out online over three rounds (Figure 1).

The stage 1 Delphi will involve asking each expert to score the importance of an outcome measure for use in weight management service outcome reporting. The scale will run from 1-9 with 1-3 indicating that the outcome is unimportant, 4-6 indicating that it is neither unimportant nor important and 7-9 indicating that it is important.

During the stage 2 Delphi, experts will be asked to score the appropriateness of outcome definitions and instruments for measurement of outcomes. Again, this will be done using a 1-9 scale with 1-3 indicating that the definition/instrument is inappropriate, 4-6 indicating that it is neither appropriate nor inappropriate and 7-9 indicating that it is appropriate.

Statistical Analysis

To assess disagreement and importance/appropriateness (and thus define consensus) the Research AND Development (RAND)/ University of California Los Angeles (UCLA) appropriateness method will be used¹⁰. This involves calculating the median score, the inter-percentile range (IPR, 30th and 70th), and the inter-percentile range adjusted for symmetry (IPRAS), for each item being rated.

1
2
3 Fitch *et al.*¹⁰ first explored using the IPR alone in an attempt to develop a method that reproduced
4 'classic' RAND definitions on panels that were multiples of 3 (which was typical in RAND's early
5 consensus studies), but could also be extended to larger panel sizes. They found that in cases when
6 agreement was good, the IPR should be narrow and in cases where there was disagreement, the IPR
7 should be wide. However, an in-depth examination of the cases of disagreement identified by the
8 IPR led to the discovery that when the ratings were symmetric, the IPR required to label an
9 indication as disagreement was smaller than when the ratings were asymmetric, with respect to the
10 middle. To overcome this, they developed the IPRAS which includes a correction factor for
11 asymmetry (*Equation 1*).

12
13
14
15
16
17
18
19
20 *Equation 1*

21
22
$$IPRAS = IPRr + (AI \times CFA)$$

23
24
25 *Where IPRr is the inter-percentile range required for disagreement when perfect symmetry exists, AI*
26 *is the asymmetry index, and CFA is the correction factor for asymmetry.*

27
28
29
30
31 The IPRAS is the threshold beyond which the IPR for a particular item indicates disagreement. Using
32 the IPRAS and the IPR to judge disagreement reproduces 'classic' RAND definitions when applied to
33 panels made up of multiples of 3, but can also be applied to panels of any size¹⁰. Variations on the
34 stringencies of definitions of disagreement exist¹¹ but similar examples of Delphi studies in health
35 services research have used the classic definition^{14;15}. In *Equation 1*, the optimal values for IPRr and
36 CFA were derived following empirical work on a 9-point scale¹⁰. Fitch *et al.* found that using values
37 of 2.35 and 1.5 best reproduced the 'classic' definitions of agreement. These values will be used in
38 this analysis. We will calculate AI as the distance between the central point of the IPR ($p30+p70/2$)
39 and the central point of the scale (i.e. 5 on a 1-9 point scale.).

40
41
42
43
44
45
46 The IPRAS threshold is dependent on the symmetry of ratings about the median. Thus, each item
47 requires a different IPRAS to be calculated. Consequently, the i^{th} indication is rated with
48 disagreement if the $IPR_i > IPRAS_i$. In previous Delphi studies some have calculated the ratio of these:
49 the disagreement index^{14;16}. If the disagreement index was less than 1.0, it indicated there was no
50 disagreement for the item in question. However, this is problematic in terms of interpretation
51 because in the case that the IPR is zero, then the ratio is zero, which can cause confusion. For this
52
53
54
55
56
57
58
59
60

1
2
3 reason we will present IPR and IPRAS values and simply comment on whether or not there is
4 disagreement (i.e. when $IPR_i > IPRAS_i$).
5

6
7 Judgement of appropriateness/importance also follows the classic RAND definitions, and this is
8 assessed simply as whether the median rating falls between 1 to 3 (inappropriate/unimportant), 4
9 and 6 (unsure), or 7 and 9 (appropriate/important).
10

11
12 At the end of each Delphi round, the median rating will be determined for individual
13 outcomes/instruments and the distribution of ratings summarised in analysis conducted by Clinvivo
14 and transferred to our research group (Figure 1).
15

16
17 During both stage 1 and stage 2, participants will be given 2 weeks to complete each round of the
18 Delphi and will be reminded of the deadline for completion before starting the process. Participants
19 will also be sent a reminder email 1 day before the deadline for each round.
20
21
22
23
24
25

26 Stage 1, Round 1 Delphi

27
28 The first Delphi study (stage 1) will be to select outcomes for inclusion in the core dataset. Full
29 instructions will be provided to the expert group prior to completion of stage 1 questionnaires.
30 Outcomes will be grouped under appropriate domains and full definitions of each outcome will be
31 provided. Participants will be asked to rate each outcome in turn using the 1-9 scale. During round 1,
32 there will be an option for adding free text outlining reasons for any given rating and also for
33 suggesting possible additional outcomes.
34
35
36
37
38
39
40

41 Analysis of Stage 1, Round 1

42
43 Additional outcomes listed by participants will be reviewed by two members of the study team
44 (RMM and JL) to ensure they represent new outcomes. All outcomes, excluding any rated
45 unimportant by consensus and including any new outcomes, will be carried forward to round 2.
46
47
48
49
50

51 Stage 1, Round 2 Delphi

52
53 In round 2, all experts will be asked to rate outcomes again. They will be shown their previous rating,
54 the median expert group rating and any free text comments in the hope of ratings reaching a
55 consensus. Experts will be asked to strongly consider the priority outcomes for weight management
56
57
58
59
60

1
2
3 reporting in this round. Additional questions will be added as to the appropriate number of items to
4 be included in the core outcome set.
5
6
7
8
9

10 11 Analysis of Stage 1, Round 2 12

13 All outcomes, excluding any rated unimportant by consensus and including any new outcomes, will
14 be carried forward to round 3.
15
16
17
18
19

20 Stage 1, Round 3 Delphi 21

22 In round 3, all experts will be asked to rate outcomes for the final time. They will be shown their
23 previous rating, the median expert group rating and any free text comments in the hope of ratings
24 reaching a consensus. Experts will be asked to strongly consider the priority outcomes for weight
25 management reporting in this round.
26
27
28
29
30
31

32 Analysis of Stage 1, Round 3 33

34 Using the consensus on the outcome set size and importance of outcomes, an outcome set will be
35 developed by the study team using the results of the Delphi.
36
37
38
39
40

41 Stage 2, Round 1 Delphi 42

43 The second Delphi study (stage 2) will be for definition/instrument selection. Selection of
44 instruments for inclusion in the stage 2 Delphi will be informed, as previously stated, by
45 results/ratings/suggestions from stage 1, systematic review and input from co-investigators (LJE and
46 SAS).
47
48
49

50 Full instructions will be provided prior to completion of stage 2 questionnaires. As per stage 1,
51 instruments will be grouped under appropriate domains and full definitions of each instrument will
52 be provided. As stated, participants will be asked to rate each instrument in turn using a 1-9 scale of
53 appropriateness (rather than importance). During the first round of the stage 2 instrument selection
54
55
56
57
58
59
60

1
2
3 process , there will be an option for adding text outlining reasons for any given rating and also for
4 suggesting possible additional instruments for measuring or defining outcomes.
5
6
7
8
9
10
11
12

13 14 Analysis of Stage 2, Round 1

15
16 Additional instruments listed by participants will be reviewed by two members of the study team
17 (RMM and JL) to ensure they represent new instruments. All instruments, excluding those rated
18 inappropriate by consensus and including any new instruments, will be carried forward to round 2.
19
20
21
22
23

24 Stage 2, Round 2 Delphi

25
26 In round 2, all experts will be asked to rate instruments again. They will be shown their previous
27 rating, the median expert group rating and any free text comments in the hope of ratings reaching a
28 consensus. Experts will be encouraged to rate instruments in a way that shows their preferences.
29
30
31
32
33

34 Analysis of Stage 2, Round 2

35
36 It may be that after round 2 an instrument set can be formed. Only those instruments related to an
37 outcome for which there is no established consensus will be carried over to round 3.
38
39
40
41
42

43 Stage 2, Round 3 Delphi

44
45 In round 3, all experts will be asked to select instruments for the final time. They will be shown their
46 previous rating, the median expert group rating and any free text comments in the hope of ratings
47 reaching a consensus. In this round they will be asked to select the most appropriate instrument for
48 each outcome in a binary format.
49
50
51
52
53

54 Analysis of Stage 2, Round 3

1
2
3 A final instrument set matched to the core outcome set will be formed based on the consensus. In
4 any areas where there is no consensus, the study team will adjudicate, taking account of free text
5 comments.
6
7
8
9
10
11
12
13
14

15 Data Storage

16
17 Participants' contact details, including email addresses and telephone numbers, and the answers
18 they provide, will only be stored by Clinvivo for the duration of the study. Clinvivo will not share
19 participants' contact details with any third parties and participants' answers will be stored
20 anonymously. Data will be encrypted before being stored on Clinvivo's server and prior to being
21 transferred to the University of Glasgow. On completion of the study, Clinvivo will destroy all data
22 after transferring it to the University of Glasgow. The University will securely store the data on
23 password access computers for a period of ten years following completion of the research project.
24
25
26
27
28
29
30
31
32
33

34 Ethics and dissemination

35
36 Ethical approval for this study has been received from the University of Glasgow College of Medical,
37 Veterinary and Life Sciences Ethics Committee. With regard to disseminating the results of our study,
38 we will submit a report to our funding body, the Chief Scientist Office of the Scottish Government
39 Health Department, and share early findings with PHE and Health Scotland. In addition, we hope to
40 communicate our results via conference presentations, peer review publication and also via our
41 institution's social media platforms.
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

- (1) National Institute for Health and Care Excellence. Weight Management: lifestyle services for overweight or obese adults (PH53). 2014.
- (2) Scottish Intercollegiate Guidelines Network. Management of Obesity: A national clinical guideline. 2010.
- (3) Loveman E, Frampton GK, Shepherd J, Picot J, Cooper K, Bryant J et al. The clinical effectiveness and cost-effectiveness of long-term weight management schemes for adults: a systematic review. *Health Technol Assess* 2011; 15(2):1-182.
- (4) Read S, Logue J. Variations in weight management services in Scotland: a national survey of weight management provision. *J Public Health (Oxf)* 2016; 38(3):e325-e335.
- (5) Public Health England. National mapping of weight management services 2015.
- (6) Public Health England. Standard Evaluation Framework for weight management interventions. 2009.
- (7) Hartmann-Boyce J, Johns D, Aveyard P, Lewis A, Jebb S, Phillips D et al. Managing overweight and obese adults: The clinical effectiveness of long-term weight management schemes for adults. 2013.
- (8) <https://www.gov.uk/government/publications/adult-weight-management-services-collect-and-record-data>. Accessed 2017.
- (9) <https://www.gov.uk/government/publications/adult-weight-management-key-performance-indicators>. Accessed 2017.

- 1
2
3 (10) Fitch K, Bernstein SJ, Aguilar DM, Burnand B, LaCalle JR, Lazaro P et al. The RAND/UCLA
4 Appropriateness Method User's Manual. 2001. RAND, Santa Monica.
5
6
7 (11) Park RE, Fink A, Brook RH, Chassin MR, Kahn KL, Merrick NJ et al. Physician ratings of
8 appropriate indications for six medical and surgical procedures. *Am J Public Health* 1986;
9 76(7):766-772.
10
11
12
13
14
15
16
17

18 **Authors' contributions**

19
20 RMM and JL drafted the protocol. LJE and SAS critically reviewed the protocol. RMM and JL finalised
21 the protocol.
22
23
24
25
26
27
28

29 **Funding Statement**

30
31 This work was supported by the Chief Scientist Office of the Scottish Government Health
32 Department, grant reference number CGA/17/08.
33
34

35 SAS was supported by a MRC Strategic Award (MC-PC-13027, MC_UU_12017_14 and SPHSU14).
36
37
38
39
40
41
42

43 **Competing interests statement**

44 JL leads a joint working project between University of Glasgow, NHS Greater Glasgow and Clyde,
45 MSD and Astra Zeneca. The project also involved an educational grant from Janssen. JL received
46 funding to attend a conference from Novo Nordisk.
47
48

49 LJE has a part time secondment with PHE.
50
51
52
53
54
55
56
57
58
59
60

Figure Legends

Figure 1. Schematic outlining the two stage Delphi study. In order to develop a core outcome set and definition/instrument set, Delphi methodology will be used to gain consensus from expert groups. Two Delphis (stage 1 and stage 2) will be carried out online over three rounds of questionnaires. The stage 1 Delphi will focus on development of a core outcome set. The stage 2 Delphi will focus on corresponding definition/instrument selection. PHE, Public Health England; SEF, standard evaluation framework; KPI, key performance indicator.

For peer review only

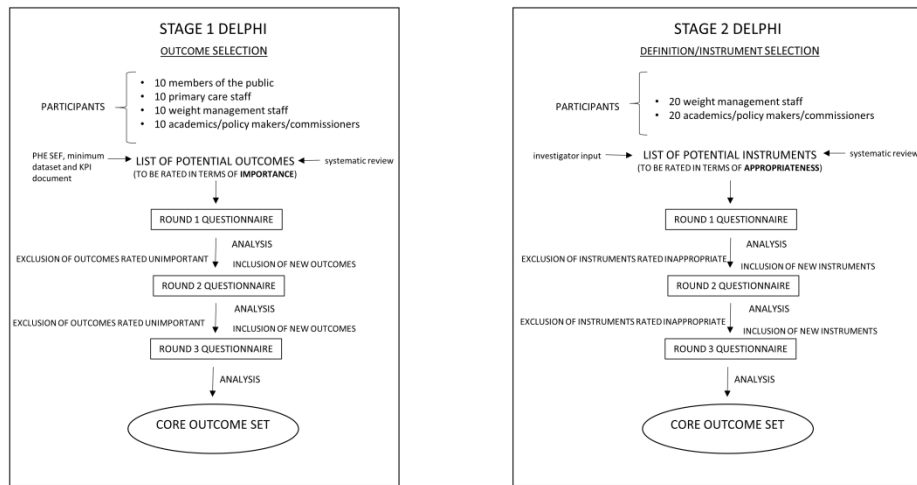


Figure 1. Schematic outlining the two stage Delphi study. In order to develop a core outcome set and definition/instrument set, Delphi methodology will be used to gain consensus from expert groups. Two Delphis (stage 1 and stage 2) will be carried out online over three rounds of questionnaires. The stage 1 Delphi will focus on development of a core outcome set. The stage 2 Delphi will focus on corresponding definition/instrument selection. PHE, Public Health England; SEF, standard evaluation framework; KPI, key performance indicator.

338x190mm (300 x 300 DPI)

BMJ Open

Development of a core outcome set for behavioural weight management programmes for adults with overweight and obesity: protocol for obtaining expert consensus using Delphi methodology

| | |
|---------------------------------|---|
| Journal: | <i>BMJ Open</i> |
| Manuscript ID | bmjopen-2018-025193.R1 |
| Article Type: | Protocol |
| Date Submitted by the Author: | 25-Oct-2018 |
| Complete List of Authors: | Mackenzie, Ruth; University of Glasgow, Institute of Cardiovascular and Medical Sciences Ells, Louisa; Teeside University, School of Health and social Care Simpson, Sharon; University of Glasgow, MRC/CSO Social and Public Health Sciences Unit Logue, Jennifer; University of Glasgow, Institute of Cardiovascular and Medical Sciences; |
| Primary Subject Heading: | Public health |
| Secondary Subject Heading: | Health policy |
| Keywords: | core outcomes, lifestyle weight management, Delphi |
| | |

SCHOLARONE™
Manuscripts

1
2
3 1 **Development of a core outcome set for behavioural weight management programmes for adults with**
4 **overweight and obesity: protocol for obtaining expert consensus using Delphi methodology**

5 2
6
7 3 *Ruth M. Mackenzie¹, Louisa J. Ells², Sharon Anne Simpson³, Jennifer Logue¹*
8

9
10 4 ¹ Institute of Cardiovascular and Medical Sciences, University of Glasgow
11

12 5 ² School of Health & Social Care, Teeside University
13

14 6 ³ Institute of Health and Wellbeing, University of Glasgow
15
16

17 7
18
19
20 8 Corresponding Author:

21
22 9 Jennifer Logue
23

24
25 10 Institute of Cardiovascular and Medical Sciences
26

27 11 University of Glasgow
28

29
30 12 126 University Place
31

32 13 Glasgow
33

34
35 14 G12 8TA
36

37 15 Jennifer.Logue@glasgow.ac.uk
38
39

40 16
41

42 17 **ABSTRACT**
43
44

45 18 **Introduction:** Weight management interventions in research studies and in clinical practice differ in
46 19 length, advice, frequency of meetings, staff, and cost. Very few real-world programmes have published
47 20 patient-related outcomes, and those that have published used different ways of reporting the
48 21 information, making it impossible to compare interventions and further develop the evidence base.
49 22 Developing a core outcome set for behavioural weight management programmes (BWMPs) for adults
50 23 with overweight and obesity will allow different BWMPs to be compared and reveal which interventions
51 24 work best for which members of the population.
52
53
54
55
56
57
58
59
60

1
2
3 25 **Methods and analysis:** An expert group, comprised of 40 people who work in, refer to, or attend BWMPs
4
5 26 for adults with overweight and obesity, will be asked to decide which outcomes services should report.
6
7 27 An online Delphi process will be employed to help the group reach consensus as to which outcomes should
8
9 28 be measured and reported, and which definitions/instruments should be utilised in order to do so. The
10
11 29 first stage of the Delphi process (3 rounds of questionnaires) will focus on outcomes while the second
12
13 30 stage (3 additional rounds of questionnaires) will focus on definition/instrument selection.

14 31 **Ethics and dissemination:** Ethical approval for this study has been received from the University of Glasgow
15
16 32 College of Medical, Veterinary and Life Sciences Ethics Committee. With regard to disseminating results,
17
18 33 a report will be submitted to our funding body, the Chief Scientist Office of the Scottish Government
19
20 34 Health Department. In addition, early findings will be shared with Public Health England (PHE) and Health
21
22 35 Scotland, and results communicated via conference presentations, peer review publication and our
23
24 36 institutions' social media platforms.

25 37 **Registration details:** The project has been registered with the COMET (Core Outcome Measures in
26
27 38 Effectiveness Trials) Initiative (<http://www.comet-initiative.org/studies/details/1056>).

29 39

31 40 **STRENGTHS AND LIMITATIONS OF THIS STUDY**

- 34 41 • The major strength of this study is that it is the first of its kind and development of a core outcome
35 42 set for BWMPs for adults with overweight and obesity is much needed in order to standardise
36 43 reporting which, in turn, will lead to a better evidence base and improvements in weight
37 44 management provision.
- 41 45 • It is a limitation that this study is wholly based in the United Kingdom (UK) as the results may need
42 46 some adaptation to be suited to real-world programmes set within other healthcare systems.
- 44 47 • However, we will use the internationally recognised Delphi method to garner opinions from a
45 48 wide range of individuals with expertise in behavioural weight management.

1
2
3 52
4
5
6 53
78 54 **INTRODUCTION**
9

10 55 Both the National Institute for Health and Care Excellence (NICE)¹ and Scottish Intercollegiate Guidelines
11 56 Network (SIGN)² guidelines outline the intervention components to be included in a community weight
12 57 management programme, namely calorie restriction, increased physical activity, and behavioural
13 58 interventions. These have proven efficacy from randomised controlled trials³. However, their
14 59 implementation in practice is inconsistent with mapping exercises in Scotland⁴ and England⁵ showing wide
15 60 variation in services in terms of inclusion criteria, referral routes, delivery format, length and cost. Few
16 61 real life services have published data and when they do publish, results can be poor with low levels of
17 62 completion and 'success', and lack of longer term outcomes.

18 63 The NICE guidance, 'Weight management: lifestyle services for overweight or obese adults'¹, identified a
19 64 number of evidence gaps. These included, reliance on studies with short follow-up, collection of data at
20 65 limited time points, small sample sizes, demographic samples that limit the ability to generalise, non-
21 66 reporting of reasons for people dropping out and lack of evidence regarding the effect of population
22 67 characteristics, such as age, gender and socio economic status, on the effectiveness of a service. They
23 68 noted a lack of comparisons between behavioural weight management programmes (BWMPs) in the
24 69 United Kingdom (UK). This lack of an evidence base means that it is not possible to issue clear guidance
25 70 as to which services are cost effective for which population groups.

26 71 Public Health England (PHE) has created a standard evaluation framework (SEF)⁶ to aid the evaluation of
27 72 real world weight management programmes. However, in their 2015 weight management mapping
28 73 exercise⁵, PHE reported that only 46% of adult weight management programmes use the SEF and, as it
29 74 simply suggests areas for reporting and potential methods of analysis, there is a huge gap in standardised
30 75 reporting. PHE had intended to analyse data from services but analysis was not possible due to the
31 76 heterogeneity of reporting which included kilograms, % weight loss, average number of completers
32 77 achieving 5% weight loss, body mass index (BMI) and more⁵. With regard to research studies, evidence
33 78 suggests similar heterogeneity in terms of the reporting of outcomes⁷.

34 79 In an attempt to address this reporting issue, PHE issued a minimum dataset⁸ which provides an important
35 80 core outcome recommendation for England, stipulating collection of certain demographics, service

1
2
3 81 details, BMI and wellbeing at baseline, on completion of the programme and at 6 months and 12 months
4
5 82 post programme. A data collection tool provides information to support the standardisation of these data
6
7 83 collection practices. This minimum dataset will be used to support PHE's recently released document on
8
9 84 adult tier 2 weight management service key performance indicators (KPIs)⁹ which provides advice as to
10
11 85 how weight status and service compliance should be reported and measured.

12
13 86 The study described herein has been funded through a Chief Scientist Office of the Scottish Government
14
15 87 Health Department grant and will serve to further validate and build upon the PHE minimum dataset⁸ and
16
17 88 KPI document⁹, while also informing a similar framework for Scotland. In addition, our research will
18
19 89 provide much needed consensus on the measurements that should be used, such as questionnaires,
20
21 90 something currently not covered in the PHE minimum dataset⁸ or KPI document⁹. Overall, this work will
22
23 91 ensure more consistency in the measurement of the effectiveness of adult weight management services,
24
25 92 leading to a better evidence base from which to identify which services are effective across a range of
26
27 93 settings.

28
29 94 Recently, a core outcome set for bariatric and metabolic surgery was successfully developed using
30
31 95 consensus methodology¹⁰. However, outcomes, including perioperative outcomes and post-operative
32
33 96 complications, are not relevant for reporting from BWMPs. Therefore, the aim of this study, which will
34
35 97 run from November 2017 until November 2018, is to gain expert consensus opinion on the core outcomes
36
37 98 that should be reported from behavioural weight management interventions for adults with overweight
38
39 99 and obesity in real world clinical practice as well as within research studies.

40
41 100 The specific study objectives are to:

- 42 101 1. Review the list of outcomes previously reported in the PHE SEF⁶, minimum dataset⁸ and KPI document⁹;
- 43 102 2. Identify additional outcomes reported in studies of structured, sustained, multi-component weight
- 44 103 management programmes for adults from a systematic review of the literature;
- 45 104 3. Select outcomes for inclusion in the core dataset using consensus methodology;
- 46 105 4. Select definitions/instruments for measuring chosen outcomes using consensus methodology.

47 106

48 107

108

109

110

111 METHODS AND ANALYSIS

112 Identification of outcomes

113 We will generate a list of outcomes by review of the PHE SEF⁶, which was itself developed from a
114 systematic review of the literature/focus groups, and from the PHE minimum dataset⁸ and KPI document⁹
115 which were developed through expert consensus and evidence from the peer review and grey literature.

116 Further outcomes will be selected by a review of included studies in the systematic review, 'The clinical
117 effectiveness of long-term weight management schemes for adults' by Hartmann-Boyce *et al.* (2013)⁷,
118 conducted during the development of NICE guidance¹. This systematic review⁷ assessed the effects of
119 multicomponent BWMPs in overweight and adults with obesity which may be applicable in the UK. To be
120 considered a multicomponent BWMP, the components of the programme had to include diet, physical
121 activity and behavioural therapy (for example, counselling sessions). The scope included commercial
122 weight loss programmes and non-commercial programmes, such as those delivered in primary care
123 settings (for example, in GP practices)⁷. It updated and expanded on an existing systematic review
124 published in 2011 by Loveman *et al.*³ and used similar methods. The Loveman systematic review³ sought
125 to assess the long-term clinical effectiveness and cost-effectiveness of
126 multicomponent weight management schemes for adults in terms of weight loss and maintenance of
127 weight loss.

128 Additional outcomes will be identified by updating the Hartmann-Boyce systematic review⁷, using the
129 same inclusion criteria but extending search dates so that studies from 1/11/2012 until 30/09/17 are
130 included. Search and selection criteria for the systematic review are identical to those of Hartmann-
131 Boyce⁷. With regard to database searches, Hartmann-Boyce⁷ searched BIOSIS, the Cochrane Database of
132 Systematic Reviews, CENTRAL, the Conference Proceedings Citation Index, the Database of Abstracts of
133 Reviews and Effects (DARE), Embase, the Health Technology Assessment database, Medline, PsychInfo,
134 and Science Citation Index for references relating to weight loss programmes. They also screened
135 references from three additional sources: reference lists in systematic reviews, documents received via
136 the NICE call for evidence, and studies excluded from Loveman³ that they wished to re-examine. Studies

1
2
3 137 selected for inclusion had to be structured, sustained, multi-component adult weight management
4
5 138 programmes with interventions which were a combination of diet and physical activity with a behaviour
6
7 139 change strategy to influence lifestyle. In addition, programmes were required to include a follow-up of
8
9 140 more than 12 months and be delivered in the health sector, in the community or commercially (i.e.
10
11 141 applicable to the NHS).

12 142 Two review authors will independently assess the abstracts of studies resulting from our literature search.
13
14 143 Full text copies of studies appearing to meet the inclusion criteria will be further independently assessed
15
16 144 by the 2 reviewers. Following discussion, agreement will be reached as to which studies to include. Any
17
18 145 new outcomes will then be identified from the selected studies from both Hartmann-Boyce⁷ and the
19
20 146 updated review.

21
22 147

23 24 148 **Identification of Instruments**

25
26 149 By review of the studies identified during the systematic reviews previously described, we will list
27
28 150 instruments and definitions for selected outcomes. The study investigators will review this list and add
29
30 151 any further suitable instruments.

31
32 152

33 34 35 153 **Data Analysis and Presentation**

36
37 154 For analysis purposes, the data will be tabulated so that the outcomes and instruments to be included in
38
39 155 our Delphi are listed and the study/studies from which they were identified are displayed. Outcomes and
40
41 156 instruments will be grouped under appropriate domains following review of selected outcomes.

42
43 157

44 45 46 158 **Patient and Public Involvement**

47
48 159 We will develop our core outcome set by means of consensus from an expert group. The sampling frame
49
50 160 will aim to include members of the public with experience of NHS, local authority or commercial adult
51
52 161 BWMPs in the UK, academics/policy makers/commissioners working in weight management, staff
53
54 162 currently involved in delivering a BWMP for adults (without significant policy involvement), and primary

1
2
3 163 care staff (referrers). Consensus methodology will ensure that the opinions and preferences of members
4
5 164 of the public will be given the same weighting as those of the other experts.
6

7 165 There is no published agreement on the optimal size of an expert group; pragmatism is required while
8
9 166 ensuring a range of opinions is garnered. Experience suggests a greater than 80% completion rate of
10
11 167 Delphi questionnaires^{10,11}. We will pre-approach potential volunteers to get agreement to participate
12
13 168 from 10 members of the public, 20 academics/policy makers/commissioners, 20 weight management staff
14
15 169 and 10 primary care staff. Forty experts will complete each of the two separate Delphi processes.

16
17 170 For the first Delphi process (stage 1, outcome selection), 10 members of the public, 10 academics/policy
18
19 171 makers/commissioners, 10 weight management staff and 10 primary care staff will be invited to
20
21 172 participate.

22
23 173 For the second Delphi (stage 2, instrument selection), 20 academics/policy makers/commissioners and 20
24
25 174 weight management staff will be invited to participate with further members recruited if any of the
26
27 175 original group (the 10 from each group who completed stage 1) have dropped out after the stage 1 Delphi.
28
29 176 The stage 2 Delphi will involve reading papers, looking at metrics and assessing validity of
30
31 177 instruments/questionnaires. With such a level of knowledge and expertise required, members of the
32
33 178 public and primary care staff will not be involved in this stage of the Delphi process.

34
35 179 A small monetary incentive (a £35 gift voucher for either John Lewis or Amazon, depending on preference)
36
37 180 will be offered to members of the public and primary care staff as this study is not of any direct benefit to
38
39 181 them and could not be considered part of their role.

40
41 182 Staff working in weight management, academics/policy makers/commissioners and primary care staff will
42
43 183 be recruited by email from the investigators and their personal contacts, and also via an email from the
44
45 184 Association for the Study of Obesity. An information letter outlining the study will be attached to emails.
46
47 185 On registering interest in our study, we will ask volunteers from these groups to provide us with
48
49 186 information as to their role and geographical location within the UK.

50
51 187 Members of the public will be recruited by email from the Association for the Study of Obesity (which has
52
53 188 lay members) and from professional contacts (a number of weight management programmes have lay
54
55 189 members on steering committees). An information letter outlining the study will be attached to emails.
56
57 190 (The information letter for the public will be written in lay language and will therefore differ slightly to the
58
59 191 information letter for the other groups.) We have also registered with the NIHR People in Research

1
2
3 192 website (<https://www.peopleinresearch.org/>) where our study will be advertised (following review to
4
5 193 ensure suitability for a lay audience). Our information letter will be available to download from this
6
7 194 website. On registering interest in our study, a 'job description' pro forma will be sent to members of the
8
9 195 public via email. They will be asked to complete this pro forma and return it to us by email. The pro forma
10 196 will provide us with information as to their gender, age, geographical location and experience of BWMPs.

11
12 197 In addition, Facebook and Twitter will be used to recruit members of the public, weight management staff,
13
14 198 academics/policy makers/commissioners and primary care staff. Facebook posts and Tweets will link to a
15
16 199 Mailchimp recruitment page where volunteers will be able to register their interest. On doing so, they will
17
18 200 receive the appropriate information letter. Weight management staff, academics/policy
19 201 makers/commissioners and primary care staff will be asked to provide us with information as to their role
20
21 202 and geographical location within the UK, and members of the public will be asked to complete the job
22
23 203 description pro forma.

24
25 204 Following provision of information regarding role and geographical location from weight management
26
27 205 staff, academics/policy makers/commissioners and primary care staff, and the return of completed pro
28
29 206 formas from members of the public, selection of volunteers to participate will commence. Selection will
30
31 207 be based on our sampling framework which is outlined below. Volunteers will be sent an email to thank
32
33 208 them for their interest and inform them if they have been selected to participate or not. A list of selected
34
35 209 volunteers' names and email addresses will then be sent to Clinvivo (www.clinvivo.com, a spin-out
36
37 210 company of the University of Warwick) who will be conducting the Delphi process. Clinvivo will then
38
39 211 contact these individuals by email, providing a link to the online Delphi questionnaire and instructions as
40
41 212 to how to complete it.

42
43 213 On completion of the study, all participants (including members of the public) will be sent (by email) a
44
45 214 copy of the final outcome and definition/instrument sets. In addition, where consent has been given,
46
47 215 participants (including members of the public) will be named as contributors in the results publication.

48
49 216

50 217 Sampling Framework

51
52 218 To ensure our volunteers are a representative UK group, of the 20 weight management staff selected, at
53
54 219 least 50% will be from England. Similarly, at least 50% of the 20 academic/policy maker/commissioner
55
56 220 group will be from England. 8 of the 20 (40%) will be academics, 6 of the 20 (30%) will be policy makers

1
2
3 221 and 6 of the 20 (30%) will be commissioners. At least 50% of the 10 primary care staff selected will also
4
5 222 be from England. With regard to members of the public, more than 50% will have experience of
6
7 223 commercial BWMPs, more than 50% will be of working age, more than 30% will be male and less than
8
9 224 30% will be from any one region of the UK.

10
11 225

12 13 226 **Delphi Survey**

14
15
16 227 In order to develop our core outcome dataset, Delphi methodology will be used to gain consensus from
17
18 228 our expert group. Two Delphis (stage 1 and stage 2) will be carried out using an online system developed
19
20 229 and conducted by Clinvivo. Each Delphi will be carried out online over three sequential rounds with the
21
22 230 same group of participants (Figure 1). For both stage 1 and stage 2 Delphis, only those who complete a
23
24 231 questionnaire in round 1 will be eligible to participate in round 2, and only those who complete round 2
25
26 232 will be eligible to participate in round 3.

27 233 The stage 1 Delphi will involve asking each expert to score the importance of an outcome measure for use
28
29 234 in weight management service outcome reporting. The scale will run from 1-9 with 1-3 indicating that the
30
31 235 outcome is unimportant, 4-6 indicating that it is neither unimportant nor important and 7-9 indicating
32
33 236 that it is important.

34 237 During the stage 2 Delphi, experts will be asked to score the appropriateness of outcome definitions and
35
36 238 instruments for measurement of outcomes. Again, this will be done using a 1-9 scale with 1-3 indicating
37
38 239 that the definition/instrument is inappropriate, 4-6 indicating that it is neither appropriate nor
39
40 240 inappropriate and 7-9 indicating that it is appropriate.

41
42 241

43 44 242 **Statistical Analysis**

45
46
47 243 To assess disagreement and importance/appropriateness (and thus define consensus) the Research AND
48
49 244 Development (RAND)/ University of California Los Angeles (UCLA) appropriateness method will be used¹¹.
50
51 245 This involves calculating the median score, the inter-percentile range (IPR, 30th and 70th), and the inter-
52
53 246 percentile range adjusted for symmetry (IPRAS), for each item being rated.

54 247 Fitch *et al.*¹¹ first explored using the IPR alone in an attempt to develop a method that reproduced 'classic'
55
56 248 RAND definitions on panels that were multiples of 3 (which was typical in RAND's early consensus studies),
57
58
59
60

1
2
3 249 but could also be extended to larger panel sizes. They found that in cases when agreement was good, the
4
5 250 IPR should be narrow and in cases where there was disagreement, the IPR should be wide. However, an
6
7 251 in-depth examination of the cases of disagreement identified by the IPR led to the discovery that when
8
9 252 the ratings were symmetric, the IPR required to label an indication as disagreement was smaller than
10
11 253 when the ratings were asymmetric, with respect to the middle. To overcome this, they developed the
12
13 254 IPRAS which includes a correction factor for asymmetry (*Equation 1*).

14 255 *Equation 1*

15
16 256
$$IPRAS = IPRr + (AI \times CFA)$$

17
18
19 257 *Where IPRr is the inter-percentile range required for disagreement when perfect symmetry exists, AI is the*
20
21 258 *asymmetry index, and CFA is the correction factor for asymmetry.*

22
23 259
24
25
26 260 The IPRAS is the threshold beyond which the IPR for a particular item indicates disagreement. Using the
27
28 261 IPRAS and the IPR to judge disagreement reproduces 'classic' RAND definitions when applied to panels
29
30 262 made up of multiples of 3, but can also be applied to panels of any size¹¹. Variations on the stringencies
31
32 263 of definitions of disagreement exist¹² but similar examples of Delphi studies in health services research
33
34 264 have used the classic definition¹³⁻¹⁸. In *Equation 1*, the optimal values for IPRr and CFA were derived
35
36 265 following empirical work on a 9-point scale¹¹. Fitch *et al.* found that using values of 2.35 and 1.5 best
37
38 266 reproduced the 'classic' definitions of agreement. These values will be used in this analysis. We will
39
40 267 calculate AI as the distance between the central point of the IPR ($p_{30}+p_{70}/2$) and the central point of the
41
42 268 scale (i.e. 5 on a 1-9 point scale.).

43
44 269 The IPRAS threshold is dependent on the symmetry of ratings about the median. Thus, each item requires
45
46 270 a different IPRAS to be calculated. Consequently, the i^{th} indication is rated with disagreement if the $IPR_i >$
47
48 271 $IPRAS_i$. In previous Delphi studies some have calculated the ratio of these: the disagreement index^{14;16;18}.
49
50 272 If the disagreement index was less than 1.0, it indicated there was no disagreement for the item in
51
52 273 question. However, this is problematic in terms of interpretation because in the case that the IPR is zero,
53
54 274 then the ratio is zero, which can cause confusion. For this reason we will present IPR and IPRAS values and
55
56 275 simply comment on whether or not there is disagreement (i.e. when $IPR_i > IPRAS_i$).

1
2
3 276 Judgement of appropriateness/importance also follows the classic RAND definitions, and this is assessed
4
5 277 simply as whether the median rating falls between 1 to 3 (inappropriate/unimportant), 4 and 6 (unsure),
6
7 278 or 7 and 9 (appropriate/important).

8
9 279 At the end of each Delphi round, the median rating will be determined for individual
10
11 280 outcomes/instruments and the distribution of ratings summarised in analysis conducted by Clinvivo and
12
13 281 transferred to our research group (Figure 1).

14
15 282 During both stage 1 and stage 2, participants will be given 2 weeks to complete each round of the Delphi
16
17 283 and will be reminded of the deadline for completion before starting the process. Participants will also be
18
19 284 sent a reminder email 1 day before the deadline for each round.

20
21 285

22 23 286 Stage 1, Round 1 Delphi

24
25
26 287 The first Delphi study (stage 1) will be to select outcomes for inclusion in the core dataset. Full instructions
27
28 288 will be provided to the expert group prior to completion of stage 1 questionnaires. Outcomes will be
29
30 289 grouped under appropriate domains (broadly based on the PHE SEF⁶ and broadly following the weight
31
32 290 management chronological pathway) and full definitions of each domain and outcome will be provided in
33
34 291 lay language. Participants will be asked to rate each outcome in turn using the 1-9 scale. During round 1,
35
36 292 there will be an option for adding free text outlining reasons for any given rating and also for suggesting
37
38 293 possible additional outcomes.

39 294

40 41 295 Analysis of Stage 1, Round 1

42
43 296 Additional outcomes listed by participants will be reviewed by two members of the study team (RMM and
44
45 297 JL) to ensure they represent new outcomes. All outcomes, excluding any rated unimportant by consensus
46
47 298 and including any new outcomes, will be carried forward to round 2.

48
49 299

50 51 52 300 Stage 1, Round 2 Delphi

53
54 301 In round 2, all experts will be asked to rate outcomes again. They will be shown their previous rating, the
55
56 302 median expert group rating and any free text comments in the hope of ratings reaching a consensus.

1
2
3 303 Experts will be asked to strongly consider the priority outcomes for weight management reporting in this
4 304 round. Additional questions will be added as to the appropriate number of items to be included in the
5 305 core outcome set.
6
7

8
9 306

10
11 307 Analysis of Stage 1, Round 2

12
13
14 308 All outcomes, excluding any rated unimportant by consensus and including any new outcomes, will be
15 309 carried forward to round 3.
16

17
18 310

19
20
21 311 Stage 1, Round 3 Delphi

22
23 312 In round 3, all experts will be asked to rate outcomes for the final time. They will be shown their previous
24 313 rating, the median expert group rating and any free text comments in the hope of ratings reaching a
25 314 consensus. Should it be the case that a large number of outcomes are being rated as important at this
26 315 stage, the need to decide which outcomes should take priority for weight management reporting will be
27 316 reinforced to experts and they will be asked to rate only these priority outcomes as important. This will
28 317 ensure development of a core outcome set of a manageable/practical size.
29

30
31
32 318 Analysis of Stage 1, Round 3

33
34
35 319 Using the consensus on the outcome set size and importance of outcomes, an outcome set will be
36 320 developed by the study team using the results of the Delphi.
37

38
39
40 321

41
42
43 322 Stage 2, Round 1 Delphi

44
45
46 323 The second Delphi study (stage 2) will be for definition/instrument selection. Selection of instruments for
47 324 inclusion in the stage 2 Delphi will be informed, as previously stated, by results/ratings/suggestions from
48 325 stage 1, systematic review and input from co-investigators (LJE and SAS).
49

50
51
52 326 Full instructions will be provided prior to completion of stage 2 questionnaires. As per stage 1, instruments
53 327 will be grouped under appropriate domains and full definitions of each instrument will be provided. As
54 328 stated, participants will be asked to rate each instrument in turn using a 1-9 scale of appropriateness
55
56
57
58
59

1
2
3 329 (rather than importance). During the first round of the stage 2 instrument selection process , there will be
4
5 330 an option for adding text outlining reasons for any given rating and also for suggesting possible additional
6
7 331 instruments for measuring or defining outcomes.
8

9 332

10
11 333 Analysis of Stage 2, Round 1

12
13
14 334 Additional instruments listed by participants will be reviewed by two members of the study team (RMM
15
16 335 and JL) to ensure they represent new instruments. All instruments, excluding those rated inappropriate
17
18 336 by consensus and including any new instruments, will be carried forward to round 2.
19

20 337

21
22 338 Stage 2, Round 2 Delphi

23
24
25 339 In round 2, all experts will be asked to rate instruments again. They will be shown their previous rating,
26
27 340 the median expert group rating and any free text comments in the hope of ratings reaching a consensus.
28
29 341 Experts will be encouraged to rate instruments in a way that shows their preferences.
30

31 342

32
33 343 Analysis of Stage 2, Round 2

34
35
36 344 It may be that after round 2 an instrument set can be formed. Only those instruments related to an
37
38 345 outcome for which there is no established consensus will be carried over to round 3.
39

40 346 Stage 2, Round 3 Delphi

41
42 347 In round 3, all experts will be asked to select instruments for the final time. They will be shown their
43
44 348 previous rating, the median expert group rating and any free text comments in the hope of ratings
45
46 349 reaching a consensus. In this round they will be asked to select the most appropriate instrument for each
47
48 350 outcome in a binary format.
49

50 351

51
52 352 Analysis of Stage 2, Round 3
53
54
55
56
57
58
59
60

1
2
3 353 A final instrument set matched to the core outcome set will be formed based on the consensus. In any
4 354 areas where there is no consensus, the study team will adjudicate, taking account of free text comments.

5
6
7 355

8
9
10 356 **Data Storage**

11
12 357 Participants' contact details, including email addresses and telephone numbers, and the answers they
13 358 provide, will only be stored by Clinivo for the duration of the study. Clinivo will not share participants'
14 359 contact details with any third parties and participants' answers will be stored anonymously. Data will be
15 360 encrypted before being stored on Clinivo's server and prior to being transferred to the University of
16 361 Glasgow. On completion of the study, Clinivo will destroy all data after transferring it to the University
17 362 of Glasgow. The University will securely store the data on password access computers for a period of ten
18 363 years following completion of the research project.

19
20
21 364

22
23
24 365

25
26
27 366

28
29
30 367

31
32
33 368

34
35
36 369

37
38
39 370

40
41
42 371

43
44
45 372 **Ethics**

46
47 373 Ethical approval for this study has been received from the University of Glasgow College of Medical,
48 374 Veterinary and Life Sciences Ethics Committee.

49
50
51 375

52
53
54 376 **Dissemination**

1
2
3 377 With regard to disseminating the results of our study, we will communicate our results via peer review
4 378 publication, conference presentations, professional societies and also via our institution's social media
5 379 platforms.

6
7
8
9 380 In addition, we will submit a report to our funding body, the Chief Scientist Office of the Scottish
10 381 Government Health Department. We will also share early findings with PHE and Health Scotland. We will
11 382 be in full discussion with both bodies to ensure that our work informs their evaluation plans for BWMPs
12 383 for adults with overweight and obesity.

13
14
15
16 384 Our study is, of course, restricted to the UK. This is due to BWMPs and their settings within health services
17 385 being fairly country-specific. For example, in France and the Netherlands there is no health insurance
18 386 funding of BWMPs and, in the USA, obesity services are tertiary, combining behavioural programmes with
19 387 medication and bariatric surgery. In addition, instruments, such as language and health economic models,
20 388 can be country-specific. However, it our belief that these differences are subtle and that, particularly for
21 389 trials, our core outcome and definition/instrument sets could be used internationally with some
22 390 adaptation.

23
24
25
26
27
28
29 391

30
31
32 392

33
34 393

35
36
37 394

38
39 395

40
41
42 396

43
44 397

45
46
47 398

48
49 399 **REFERENCES**

50
51 400

52
53 401 (1) National Institute for Health and Care Excellence. Weight Management: lifestyle services for
54 402 overweight or obese adults (PH53). 2014.

55
56 403

- 1
2
3 404 (2) Scottish Intercollegiate Guidelines Network. Management of Obesity: A national clinical guideline.
4 405 2010.
5 406
6
7 407 (3) Loveman E, Frampton GK, Shepherd J, Picot J, Cooper K, Bryant J et al. The clinical effectiveness
8 408 and cost-effectiveness of long-term weight management schemes for adults: a systematic review.
9 409 *Health Technol Assess* 2011; 15(2):1-182.
10
11
12 410 (4) Read S, Logue J. Variations in weight management services in Scotland: a national survey of weight
13 411 management provision. *J Public Health (Oxf)* 2016; 38(3):e325-e335.
14
15 412 (5) Public Health England. National mapping of weight management services 2015. 2018.
16 413
17
18 414 (6) Public Health England. Standard Evaluation Framework for weight management interventions.
19 415 2009.
20 416
21
22 417 (7) Hartmann-Boyce J, Johns D, Aveyard P, Lewis A, Jebb S, Phillips D et al. Managing overweight and
23 418 obese adults: The clinical effectiveness of long-term weight management schemes for adults.
24 419 2013.
25 420
26
27
28 421 (8) [https://www.gov.uk/government/publications/adult-weight-management-services-collect-and-](https://www.gov.uk/government/publications/adult-weight-management-services-collect-and-record-data)
29 422 [record-data](https://www.gov.uk/government/publications/adult-weight-management-services-collect-and-record-data). 2018.
30 423
31
32 424 (9) [https://www.gov.uk/government/publications/adult-weight-management-key-performance-](https://www.gov.uk/government/publications/adult-weight-management-key-performance-indicators)
33 425 [indicators](https://www.gov.uk/government/publications/adult-weight-management-key-performance-indicators). 2018.
34 426
35
36
37 427 (10) Coulman KD, Hopkins J, Brookes ST, Chalmers K, Main B, Owen-Smith A et al. A Core Outcome Set
38 428 for the Benefits and Adverse Events of Bariatric and Metabolic Surgery: The BARIACT Project. *PLoS*
39 429 *Med* 2016; 13(11):e1002187.
40
41 430 (11) Fitch K, Bernstein SJ, Aguilar DM, Burnand B, LaCalle JR, Lazaro P et al. The RAND/UCLA
42 431 Appropriateness Method User's Manual. 2001. RAND, Santa Monica.
43 432
44
45 433 (12) Park RE, Fink A, Brook RH, Chassin MR, Kahn KL, Merrick NJ et al. Physician ratings of appropriate
46 434 indications for six medical and surgical procedures. *Am J Public Health* 1986; 76(7):766-772.
47
48
49 435 (13) Eldridge SM, Lancaster GA, Campbell MJ, Thabane L, Hopewell S, Coleman CL et al. Defining
50 436 Feasibility and Pilot Studies in Preparation for Randomised Controlled Trials: Development of a
51 437 Conceptual Framework. *PLoS One* 2016; 11(3):e0150205.
52
53 438 (14) Froud R, Eldridge S, Kovacs F, Breen A, Bolton J, Dunn K et al. Reporting outcomes of back pain
54 439 trials: a modified Delphi study. *Eur J Pain* 2011; 15(10):1068-1074.
55
56
57
58
59
60

- 1
2
3 440 (15) Petrou S, Rivero-Arias O, Dakin H, Longworth L, Oppe M, Froud R et al. Preferred reporting items
4 441 for studies mapping onto preference-based outcome measures: The MAPS statement. *J Med Econ*
5 442 2015; 18(11):851-857.
6
7 443 (16) Pincus T, Miles C, Froud R, Underwood M, Carnes D, Taylor SJ. Methodological criteria for the
8 444 assessment of moderators in systematic reviews of randomised controlled trials: a consensus
9 445 study. *BMC Med Res Methodol* 2011; 11:14.
10
11 446 (17) Pinnock H, Barwick M, Carpenter CR, Eldridge S, Grandes G, Griffiths CJ et al. Standards for
12 447 Reporting Implementation Studies (StaRI) Statement. *BMJ* 2017; 356:i6795.
13
14 448 (18) Taylor WJ, Schumacher HR, Jr., Baraf HS, Chapman P, Stamp L, Doherty M et al. A modified Delphi
15 449 exercise to determine the extent of consensus with OMERACT outcome domains for studies of
16 450 acute and chronic gout. *Ann Rheum Dis* 2008; 67(6):888-891.
17 451
18 452
19
20
21
22 453
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

454 **AUTHORS' CONTRIBUTIONS**

455 RMM and JL drafted the protocol. LJE and SAS critically reviewed the protocol. RMM and JL finalised the
456 protocol.

457

458

459 **FUNDING STATEMENT**

460 This work was supported by the Chief Scientist Office of the Scottish Government Health Department,
461 grant reference number CGA/17/08.

462 SAS was supported by a MRC Strategic Award (MC-PC-13027, MC_UU_12017_14 and SPHSU14).

463

464

465 **COMPETING INTERESTS STATEMENT**

466 JL leads a joint working project between University of Glasgow, NHS Greater Glasgow and Clyde, MSD and
467 Astra Zeneca. The project also involved an educational grant from Janssen. JL received funding to attend
468 a conference from Novo Nordisk.

469 LJE has a part time secondment with PHE.

470

471

472 **FIGURE LEGENDS**

473 **Figure 1. Schematic outlining the two stage Delphi study.** In order to develop a core outcome set and
474 definition/instrument set, Delphi methodology will be used to gain consensus from expert groups. Two
475 Delphis (stage 1 and stage 2) will be carried out online over three rounds of questionnaires. The stage 1
476 Delphi will focus on development of a core outcome set. The stage 2 Delphi will focus on corresponding
477 definition/instrument selection. PHE, Public Health England; SEF, standard evaluation framework; KPI, key
478 performance indicator.

1
2
3 479
4
5
6 480
7
8 481
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

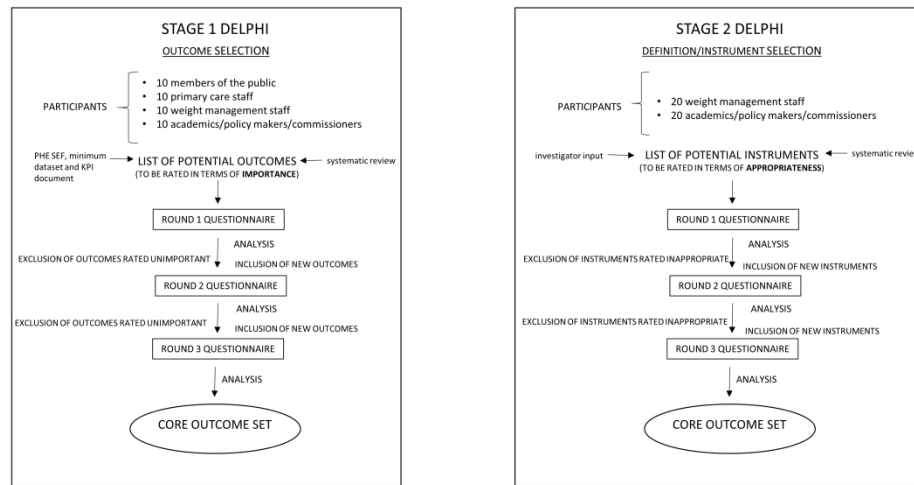


Figure 1. Schematic outlining the two stage Delphi study. In order to develop a core outcome set and definition/instrument set, Delphi methodology will be used to gain consensus from expert groups. Two Delphis (stage 1 and stage 2) will be carried out online over three rounds of questionnaires. The stage 1 Delphi will focus on development of a core outcome set. The stage 2 Delphi will focus on corresponding definition/instrument selection. PHE, Public Health England; SEF, standard evaluation framework; KPI, key performance indicator.

338x190mm (300 x 300 DPI)

BMJ Open

Development of a core outcome set for behavioural weight management programmes for adults with overweight and obesity: protocol for obtaining expert consensus using Delphi methodology

| | |
|---------------------------------|---|
| Journal: | <i>BMJ Open</i> |
| Manuscript ID | bmjopen-2018-025193.R2 |
| Article Type: | Protocol |
| Date Submitted by the Author: | 12-Dec-2018 |
| Complete List of Authors: | Mackenzie, Ruth; University of Glasgow, Institute of Cardiovascular and Medical Sciences Ells, Louisa; Teeside University, School of Health and social Care Simpson, Sharon; University of Glasgow, MRC/CSO Social and Public Health Sciences Unit Logue, Jennifer; University of Glasgow, Institute of Cardiovascular and Medical Sciences; |
| Primary Subject Heading: | Public health |
| Secondary Subject Heading: | Health policy |
| Keywords: | core outcomes, lifestyle weight management, Delphi |
| | |

SCHOLARONE™
Manuscripts

1
2
3 1 **Development of a core outcome set for behavioural weight management programmes for adults**
4 **with overweight and obesity: protocol for obtaining expert consensus using Delphi methodology**

5 2
6
7 3 *Ruth M. Mackenzie¹, Louisa J. Ells², Sharon Anne Simpson³, Jennifer Logue¹*
8

9
10 4 ¹Institute of Cardiovascular and Medical Sciences, University of Glasgow
11

12 5 ²School of Health & Social Care, Teeside University
13

14 6 ³Institute of Health and Wellbeing, University of Glasgow
15
16

17 7
18
19
20 8 Corresponding Author:
21

22 9 Jennifer Logue
23

24
25 10 Institute of Cardiovascular and Medical Sciences
26

27 11 University of Glasgow
28

29
30 12 126 University Place
31

32 13 Glasgow
33

34
35 14 G12 8TA
36

37 15 Jennifer.Logue@glasgow.ac.uk
38
39

40 16
41

42 17 **ABSTRACT**
43

44
45 18 **Introduction:** Weight management interventions in research studies and in clinical practice differ in
46 19 length, advice, frequency of meetings, staff, and cost. Very few real-world programmes have
47 20 published patient-related outcomes, and those that have published used different ways of reporting
48 21 the information, making it impossible to compare interventions and further develop the evidence
49 22 base. Developing a core outcome set for behavioural weight management programmes (BWMPs) for
50 23 adults with overweight and obesity will allow different BWMPs to be compared and reveal which
51 24 interventions work best for which members of the population.
52
53
54
55

56
57 25 **Methods and analysis:** An expert group, comprised of 40 people who work in, refer to, or attend
58 26 BWMPs for adults with overweight and obesity, will be asked to decide which outcomes services
59
60

1
2
3 27 should report. An online Delphi process will be employed to help the group reach consensus as to
4
5 28 which outcomes should be measured and reported, and which definitions/instruments should be
6
7 29 utilised in order to do so. The first stage of the Delphi process (3 rounds of questionnaires) will focus
8
9 30 on outcomes while the second stage (3 additional rounds of questionnaires) will focus on
10
11 31 definition/instrument selection.

12
13 32 **Ethics and dissemination:** Ethical approval for this study has been received from the University of
14
15 33 Glasgow College of Medical, Veterinary and Life Sciences Ethics Committee. With regard to
16
17 34 disseminating results, a report will be submitted to our funding body, the Chief Scientist Office of the
18
19 35 Scottish Government Health Department. In addition, early findings will be shared with Public Health
20
21 36 England (PHE) and Health Scotland, and results communicated via conference presentations, peer
22
23 37 review publication and our institutions' social media platforms.

24
25 38 **Registration details:** The project has been registered with the COMET (Core Outcome Measures in
26
27 39 Effectiveness Trials) Initiative (<http://www.comet-initiative.org/studies/details/1056>).

28
29
30
31
32
33
34
35
36
37
38
39
40

41 **STRENGTHS AND LIMITATIONS OF THIS STUDY**

- 42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- The major strength of this study is that it is the first of its kind and development of a core outcome set for BWMPs for adults with overweight and obesity is much needed in order to standardise reporting which, in turn, will lead to a better evidence base and improvements in weight management provision.
 - It is a limitation that this study is wholly based in the United Kingdom (UK) as the results may need some adaptation to be suited to real-world programmes set within other healthcare systems.
 - The recognised method for core outcome set development, the Delphi method, will be used to garner opinions from a wide range of individuals with expertise in behavioural weight management.
 - Review of all existing qualitative research studies will not be undertaken when generating the initial list of outcomes. However, qualitative work will be performed during core outcome set development as part of the Delphi process.

58

59 INTRODUCTION

60 Both the National Institute for Health and Care Excellence (NICE)¹ and Scottish Intercollegiate
61 Guidelines Network (SIGN)² guidelines outline the intervention components to be included in a
62 community weight management programme, namely calorie restriction, increased physical activity,
63 and behavioural interventions. These have proven efficacy from randomised controlled trials³.
64 However, their implementation in practice is inconsistent with mapping exercises in Scotland⁴ and
65 England⁵ showing wide variation in services in terms of inclusion criteria, referral routes, delivery
66 format, length and cost. Few real life services have published data and when they do publish, results
67 can be poor with low levels of completion and 'success', and lack of longer term outcomes.

68 The NICE guidance, 'Weight management: lifestyle services for overweight or obese adults'¹, identified
69 a number of evidence gaps. These included, reliance on studies with short follow-up, collection of data
70 at limited time points, small sample sizes, demographic samples that limit the ability to generalise,
71 non-reporting of reasons for people dropping out and lack of evidence regarding the effect of
72 population characteristics, such as age, gender and socio economic status, on the effectiveness of a
73 service. They noted a lack of comparisons between behavioural weight management programmes
74 (BWMPs) in the United Kingdom (UK). This lack of an evidence base means that it is not possible to
75 issue clear guidance as to which services are cost effective for which population groups.

76 Public Health England (PHE) has created a standard evaluation framework (SEF)⁶ to aid the evaluation
77 of real world weight management programmes. However, in their 2015 weight management mapping
78 exercise⁵, PHE reported that only 46% of adult weight management programmes use the SEF and, as
79 it simply suggests areas for reporting and potential methods of analysis, there is a huge gap in
80 standardised reporting. PHE had intended to analyse data from services but analysis was not possible
81 due to the heterogeneity of reporting which included kilograms, % weight loss, average number of
82 completers achieving 5% weight loss, body mass index (BMI) and more⁵. With regard to research
83 studies, evidence suggests similar heterogeneity in terms of the reporting of outcomes⁷.

84 In an attempt to address this reporting issue, PHE issued a minimum dataset⁸ which provides an
85 important core outcome recommendation for England, stipulating collection of certain demographics,
86 service details, BMI and wellbeing at baseline, on completion of the programme and at 6 months and
87 12 months post programme. A data collection tool provides information to support the
88 standardisation of these data collection practices. This minimum dataset will be used to support PHE's
89 recently released document on adult tier 2 weight management service key performance indicators

1
2
3 90 (KPIs)⁹ which provides advice as to how weight status and service compliance should be reported and
4
5 91 measured.

6
7 92 The study described herein has been funded through a Chief Scientist Office of the Scottish
8
9 93 Government Health Department grant and will serve to further validate and build upon the PHE
10
11 94 minimum dataset⁸ and KPI document⁹, while also informing a similar framework for Scotland. In
12
13 95 addition, our research will provide much needed consensus on the measurements that should be
14
15 96 used, such as questionnaires, something currently not covered in the PHE minimum dataset⁸ or KPI
16
17 97 document⁹. Overall, this work will ensure more consistency in the measurement of the effectiveness
18
19 98 of adult weight management services, leading to a better evidence base from which to identify which
20
21 99 services are effective across a range of settings.

22 100 Recently, a core outcome set for bariatric and metabolic surgery was successfully developed using
23
24 101 consensus methodology¹⁰. However, outcomes, including perioperative outcomes and post-operative
25
26 102 complications, are not relevant for reporting from BWMPs. Therefore, the aim of this study, which will
27
28 103 run from November 2017 until November 2018, is to gain expert consensus opinion on the core
29
30 104 outcomes that should be reported from behavioural weight management interventions for adults with
31
32 105 overweight and obesity in real world clinical practice as well as within research studies.

33 106 The specific study objectives are to:

- 34
35 107 1. Review the list of outcomes previously reported in the PHE SEF⁶, minimum dataset⁸ and KPI
36
37 108 document⁹;
- 38
39 109 2. Identify additional outcomes reported in studies of structured, sustained, multi-component weight
40
41 110 management programmes for adults from a systematic review of the literature;
- 42
43 111 3. Select outcomes for inclusion in the core dataset using consensus methodology;
- 44
45
46 112 4. Select definitions/instruments for measuring chosen outcomes using consensus methodology.

47
48 113

49
50 114

51
52 115

53
54 116

55
56 117

57
58
59
60

118 **METHODS AND ANALYSIS**

119 **Identification of outcomes**

120 We will generate a list of outcomes by review of the PHE SEF⁶, which was itself developed from a
121 systematic review of the literature/focus groups, and from the PHE minimum dataset⁸ and KPI
122 document⁹ which were developed through expert consensus and evidence from the peer review and
123 grey literature.

124 Further outcomes will be selected by a review of included studies in the systematic review, 'The clinical
125 effectiveness of long-term weight management schemes for adults' by Hartmann-Boyce *et al.* (2013)⁷,
126 conducted during the development of NICE guidance¹. This systematic review⁷ assessed the effects
127 of multicomponent BWMPs in overweight and adults with obesity which may be applicable in the UK.
128 To be considered a multicomponent BWMP, the components of the programme had to include diet,
129 physical activity and behavioural therapy (for example, counselling sessions). The scope included
130 commercial weight loss programmes and non-commercial programmes, such as those delivered in
131 primary care settings (for example, in GP practices)⁷. It updated and expanded on an existing
132 systematic review published in 2011 by Loveman *et al.*³ and used similar methods. The Loveman
133 systematic review³ sought to assess the long-term clinical effectiveness and cost-effectiveness of
134 multicomponent weight management schemes for adults in terms of weight loss and maintenance of
135 weight loss.

136 Additional outcomes will be identified by updating the Hartmann-Boyce systematic review⁷, using
137 the same inclusion criteria but extending search dates so that studies from 1/11/2012 until 30/09/17
138 are included. Search and selection criteria for the systematic review are identical to those of
139 Hartmann-Boyce⁷. With regard to database searches, Hartmann-Boyce⁷ searched BIOSIS, the
140 Cochrane Database of Systematic Reviews, CENTRAL, the Conference Proceedings Citation Index, the
141 Database of Abstracts of Reviews and Effects (DARE), Embase, the Health Technology Assessment
142 database, Medline, PsychInfo, and Science Citation Index for references relating to weight loss
143 programmes. They also screened references from three additional sources: reference lists in
144 systematic reviews, documents received via the NICE call for evidence, and studies excluded from
145 Loveman³ that they wished to re-examine. Studies selected for inclusion had to be structured,
146 sustained, multi-component adult weight management programmes with interventions which were
147 a combination of diet and physical activity with a behaviour change strategy to influence lifestyle. In
148 addition, programmes were required to include a follow-up of more than 12 months and be
149 delivered in the health sector, in the community or commercially (i.e. applicable to the NHS).

1
2
3 150 Two review authors will independently assess the abstracts of studies resulting from our literature
4
5 151 search. Full text copies of studies appearing to meet the inclusion criteria will be further independently
6
7 152 assessed by the 2 reviewers. Following discussion, agreement will be reached as to which studies to
8
9 153 include. Any new outcomes will then be identified from the selected studies from both Hartmann-
10 154 Boyce⁷ and the updated review.
11

12 155

15 156 **Identification of Instruments**

16
17 157 By review of the studies identified during the systematic reviews previously described, we will list
18
19 158 instruments and definitions for selected outcomes. The study investigators will review this list and add
20
21 159 any further suitable instruments.
22

23 160

26 161 **Data Analysis and Presentation**

27
28 162 For analysis purposes, the data will be tabulated so that the outcomes and instruments to be included
29
30 163 in our Delphi are listed and the study/studies from which they were identified are displayed. Outcomes
31
32 164 and instruments will be grouped under appropriate domains following review of selected outcomes.
33

34 165

36 166 **Patient and Public Involvement**

37
38
39 167 We will develop our core outcome set by means of consensus from an expert group. The sampling
40
41 168 frame will aim to include members of the public with experience of NHS, local authority or commercial
42
43 169 adult BWMPs in the UK, academics/policy makers/commissioners working in weight management,
44
45 170 staff currently involved in delivering a BWMP for adults (without significant policy involvement), and
46
47 171 primary care staff (referrers). Consensus methodology will ensure that the opinions and preferences
48
49 172 of members of the public will be given the same weighting as those of the other experts.

50 173 There is no published agreement on the optimal size of an expert group; pragmatism is required while
51
52 174 ensuring a range of opinions is garnered. Experience suggests a greater than 80% completion rate of
53
54 175 Delphi questionnaires^{10;11}. We will pre-approach potential volunteers to get agreement to participate
55
56 176 from 10 members of the public, 20 academics/policy makers/commissioners, 20 weight management
57
58 177 staff and 10 primary care staff. Forty experts will complete each of the two separate Delphi processes.
59
60

1
2
3 178 For the first Delphi process (stage 1, outcome selection), 10 members of the public, 10
4 179 academics/policy makers/commissioners, 10 weight management staff and 10 primary care staff will
5 180 be invited to participate.

6
7
8
9 181 For the second Delphi (stage 2, instrument selection), 20 academics/policy makers/commissioners and
10 182 20 weight management staff will be invited to participate with further members recruited if any of the
11 183 original group (the 10 from each group who completed stage 1) have dropped out after the stage 1
12 184 Delphi. The stage 2 Delphi will involve reading papers, looking at metrics and assessing validity of
13 185 instruments/questionnaires. As in depth knowledge of academic literature and reporting tools is
14 186 required, this stage of the Delphi process will be restricted to academics/policy
15 187 makers/commissioners and weight management staff.

16
17 188 A small monetary incentive (a £35 gift voucher for either John Lewis or Amazon, depending on
18 189 preference) will be offered to members of the public and primary care staff as this study is not of any
19 190 direct benefit to them and could not be considered part of their role.

20
21
22 191 Staff working in weight management, academics/policy makers/commissioners and primary care staff
23 192 will be recruited by email from the investigators and their personal contacts, and also via an email
24 193 from the Association for the Study of Obesity. An information letter outlining the study will be
25 194 attached to emails. On registering interest in our study, we will ask volunteers from these groups to
26 195 provide us with information as to their role and geographical location within the UK.

27
28 196 Members of the public will be recruited by email from the Association for the Study of Obesity (which
29 197 has lay members) and from professional contacts (a number of weight management programmes have
30 198 lay members on steering committees). An information letter outlining the study will be attached to
31 199 emails. (The information letter for the public will be written in lay language and will therefore differ
32 200 slightly to the information letter for the other groups.) We have also registered with the NIHR People
33 201 in Research website (<https://www.peopleinresearch.org/>) where our study will be advertised
34 202 (following review to ensure suitability for a lay audience). Our information letter will be available to
35 203 download from this website. On registering interest in our study, a 'job description' pro forma will be
36 204 sent to members of the public via email. They will be asked to complete this pro forma and return it
37 205 to us by email. The pro forma will provide us with information as to their gender, age, geographical
38 206 location and experience of BWMPs.

39
40
41 207 In addition, Facebook and Twitter will be used to recruit members of the public, weight management
42 208 staff, academics/policy makers/commissioners and primary care staff. Facebook posts and Tweets will
43 209 link to a Mailchimp recruitment page where volunteers will be able to register their interest. On doing
44
45
46
47
48
49
50
51
52
53
54
55

1
2
3 210 so, they will receive the appropriate information letter. Weight management staff, academics/policy
4 211 makers/commissioners and primary care staff will be asked to provide us with information as to their
5 212 role and geographical location within the UK, and members of the public will be asked to complete
6 213 the job description pro forma.

7
8
9
10 214 Following provision of information regarding role and geographical location from weight management
11 215 staff, academics/policy makers/commissioners and primary care staff, and the return of completed
12 216 pro formas from members of the public, selection of volunteers to participate will commence.
13 217 Selection will be based on our sampling framework which is outlined below. Volunteers will be sent
14 218 an email to thank them for their interest and inform them if they have been selected to participate or
15 219 not. A list of selected volunteers' names and email addresses will then be sent to Clinivo
16 220 (www.clinivo.com, a spin-out company of the University of Warwick) who will be conducting the
17 221 Delphi process. Clinivo will then contact these individuals by email, providing a link to the online
18 222 Delphi questionnaire and instructions as to how to complete it.

19 223 On completion of the study, all participants (including members of the public) will be sent (by email)
20 224 a copy of the final outcome and definition/instrument sets. In addition, where consent has been given,
21 225 participants (including members of the public) will be named as contributors in the results publication.

22 226

23 227 Sampling Framework

24 228 To ensure our volunteers are a representative UK group, of the 20 weight management staff selected,
25 229 at least 50% will be from England. Similarly, at least 50% of the 20 academic/policy
26 230 maker/commissioner group will be from England. 8 of the 20 (40%) will be academics, 6 of the 20
27 231 (30%) will be policy makers and 6 of the 20 (30%) will be commissioners. At least 50% of the 10 primary
28 232 care staff selected will also be from England. With regard to members of the public, more than 50%
29 233 will have experience of commercial BWMPs, more than 50% will be of working age, more than 30%
30 234 will be male and less than 30% will be from any one region of the UK.

31 235

32 236 Delphi Survey

33
34
35 237 In order to develop our core outcome dataset, Delphi methodology will be used to gain consensus
36 238 from our expert group. Two Delphis (stage 1 and stage 2) will be carried out using an online system
37 239 developed and conducted by Clinivo. Each Delphi will be carried out online over three sequential
38 240 rounds with the same group of participants (Figure 1). For both stage 1 and stage 2 Delphis, only those

241 who complete a questionnaire in round 1 will be eligible to participate in round 2, and only those who
242 complete round 2 will be eligible to participate in round 3.

243 The stage 1 Delphi will involve asking each expert to score the importance of an outcome measure for
244 use in weight management service outcome reporting. The scale will run from 1-9 with 1-3 indicating
245 that the outcome is unimportant, 4-6 indicating that it is neither unimportant nor important and 7-9
246 indicating that it is important.

247 During the stage 2 Delphi, experts will be asked to score the appropriateness of outcome definitions
248 and instruments for measurement of outcomes. Again, this will be done using a 1-9 scale with 1-3
249 indicating that the definition/instrument is inappropriate, 4-6 indicating that it is neither appropriate
250 nor inappropriate and 7-9 indicating that it is appropriate.

251

252 Statistical Analysis

253 To assess disagreement and importance/appropriateness (and thus define consensus) the Research
254 ANd Development (RAND)/ University of California Los Angeles (UCLA) appropriateness method will
255 be used¹¹. This involves calculating the median score, the inter-percentile range (IPR, 30th and 70th),
256 and the inter-percentile range adjusted for symmetry (IPRAS), for each item being rated.

257 Fitch *et al.*¹¹ first explored using the IPR alone in an attempt to develop a method that reproduced
258 'classic' RAND definitions on panels that were multiples of 3 (which was typical in RAND's early
259 consensus studies), but could also be extended to larger panel sizes. They found that in cases when
260 agreement was good, the IPR should be narrow and in cases where there was disagreement, the IPR
261 should be wide. However, an in-depth examination of the cases of disagreement identified by the IPR
262 led to the discovery that when the ratings were symmetric, the IPR required to label an indication as
263 disagreement was smaller than when the ratings were asymmetric, with respect to the middle. To
264 overcome this, they developed the IPRAS which includes a correction factor for asymmetry (*Equation*
265 *1*).

266 Equation 1

$$267 \text{IPRAS} = \text{IPRr} + (\text{AI} \times \text{CFA})$$

268 *Where IPRr is the inter-percentile range required for disagreement when perfect symmetry exists, AI is*
269 *the asymmetry index, and CFA is the correction factor for asymmetry.*

270

1
2
3 271 The IPRAS is the threshold beyond which the IPR for a particular item indicates disagreement. Using
4
5 272 the IPRAS and the IPR to judge disagreement reproduces 'classic' RAND definitions when applied to
6
7 273 panels made up of multiples of 3, but can also be applied to panels of any size¹¹. Variations on the
8
9 274 stringencies of definitions of disagreement exist¹² but similar examples of Delphi studies in health
10
11 275 services research have used the classic definition¹³⁻¹⁸. In *Equation 1*, the optimal values for IPRr and
12
13 276 CFA were derived following empirical work on a 9-point scale¹¹. Fitch *et al.* found that using values of
14
15 277 2.35 and 1.5 best reproduced the 'classic' definitions of agreement. These values will be used in this
16
17 278 analysis. We will calculate AI as the distance between the central point of the IPR (p30+p70/2) and the
18
19 279 central point of the scale (i.e. 5 on a 1-9 point scale.).

20
21 280 The IPRAS threshold is dependent on the symmetry of ratings about the median. Thus, each item
22
23 281 requires a different IPRAS to be calculated. Consequently, the i^{th} indication is rated with disagreement
24
25 282 if the $IPR_i > IPRAS_i$. In previous Delphi studies some have calculated the ratio of these: the
26
27 283 disagreement index^{14;16;18}. If the disagreement index was less than 1.0, it indicated there was no
28
29 284 disagreement for the item in question. However, this is problematic in terms of interpretation because
30
31 285 in the case that the IPR is zero, then the ratio is zero, which can cause confusion. For this reason we
32
33 286 will present IPR and IPRAS values and simply comment on whether or not there is disagreement (i.e.
34
35 287 when $IPR_i > IPRAS_i$).

36
37 288 Judgement of appropriateness/importance also follows the classic RAND definitions, and this is
38
39 289 assessed simply as whether the median rating falls between 1 to 3 (inappropriate/unimportant), 4 and
40
41 290 6 (unsure), or 7 and 9 (appropriate/important).

42
43 291 At the end of each Delphi round, the median rating will be determined for individual
44
45 292 outcomes/instruments and the distribution of ratings summarised in analysis conducted by Clinivo
46
47 293 and transferred to our research group (Figure 1).

48
49 294 During both stage 1 and stage 2, participants will be given 2 weeks to complete each round of the
50
51 295 Delphi and will be reminded of the deadline for completion before starting the process. Participants
52
53 296 will also be sent a reminder email 1 day before the deadline for each round.

54
55 297

56 298 Stage 1, Round 1 Delphi

57
58 299 The first Delphi study (stage 1) will be to select outcomes for inclusion in the core dataset. Full
59
60 300 instructions will be provided to the expert group prior to completion of stage 1 questionnaires.
301
302 301 Outcomes will be grouped under appropriate domains (broadly based on the PHE SEF⁶ and broadly

1
2
3 302 following the weight management chronological pathway) and full definitions of each domain and
4
5 303 outcome will be provided in lay language. Participants will be asked to rate each outcome in turn using
6
7 304 the 1-9 scale. During round 1, there will be an option for adding free text outlining reasons for any
8
9 305 given rating and also for suggesting possible additional outcomes.

10
11 306

12 13 307 Analysis of Stage 1, Round 1

14
15 308 Additional outcomes listed by participants will be reviewed by two members of the study team (RMM
16
17 309 and JL) to ensure they represent new outcomes. All outcomes, excluding any rated unimportant by
18
19 310 consensus and including any new outcomes, will be carried forward to round 2.

20
21 311

22 23 24 312 Stage 1, Round 2 Delphi

25
26 313 In round 2, all experts will be asked to rate outcomes again. They will be shown their previous rating,
27
28 314 the median expert group rating and any free text comments in the hope of ratings reaching a
29
30 315 consensus. Experts will be asked to strongly consider the priority outcomes for weight management
31
32 316 reporting in this round. Additional questions will be added as to the appropriate number of items to
33
34 317 be included in the core outcome set.

35
36 318

37 38 319 Analysis of Stage 1, Round 2

39
40 320 All outcomes, excluding any rated unimportant by consensus and including any new outcomes, will be
41
42 321 carried forward to round 3.

43
44 322

45 46 47 323 Stage 1, Round 3 Delphi

48
49 324 In round 3, all experts will be asked to rate outcomes for the final time. They will be shown their
50
51 325 previous rating, the median expert group rating and any free text comments in the hope of ratings
52
53 326 reaching a consensus. Should it be the case that a large number of outcomes are being rated as
54
55 327 important at this stage, the need to decide which outcomes should take priority for weight
56
57 328 management reporting will be reinforced to experts and they will be asked to rate only these priority
58
59 329 outcomes as important. This will ensure development of a core outcome set of a manageable/practical
60
330 size.

1
2
3 331 Analysis of Stage 1, Round 3
4

5
6 332 Using the consensus on the outcome set size and importance of outcomes, an outcome set will be
7 333 developed by the study team using the results of the Delphi.
8
9

10 334

11
12 335 Stage 2, Round 1 Delphi
13

14
15 336 The second Delphi study (stage 2) will be for definition/instrument selection. Selection of instruments
16 337 for inclusion in the stage 2 Delphi will be informed, as previously stated, by results/ratings/suggestions
17 338 from stage 1, systematic review and input from co-investigators (LJE and SAS).
18
19

20
21 339 Full instructions will be provided prior to completion of stage 2 questionnaires. As per stage 1,
22 340 instruments will be grouped under appropriate domains and full definitions of each instrument will
23 341 be provided. As stated, participants will be asked to rate each instrument in turn using a 1-9 scale of
24 342 appropriateness (rather than importance). During the first round of the stage 2 instrument selection
25 343 process , there will be an option for adding text outlining reasons for any given rating and also for
26 344 suggesting possible additional instruments for measuring or defining outcomes.
27
28
29

30 345

31
32
33
34 346 Analysis of Stage 2, Round 1
35

36
37 347 Additional instruments listed by participants will be reviewed by two members of the study team
38 348 (RMM and JL) to ensure they represent new instruments. All instruments, excluding those rated
39 349 inappropriate by consensus and including any new instruments, will be carried forward to round 2.
40
41

42 350

43
44
45 351 Stage 2, Round 2 Delphi
46

47
48 352 In round 2, all experts will be asked to rate instruments again. They will be shown their previous rating,
49 353 the median expert group rating and any free text comments in the hope of ratings reaching a
50 354 consensus. Experts will be encouraged to rate instruments in a way that shows their preferences.
51
52

53 355

54
55
56 356 Analysis of Stage 2, Round 2
57

58
59 357 It may be that after round 2 an instrument set can be formed. Only those instruments related to an
60 358 outcome for which there is no established consensus will be carried over to round 3.

1
2
3 359 Stage 2, Round 3 Delphi
4

5
6 360 In round 3, all experts will be asked to select instruments for the final time. They will be shown their
7
8 361 previous rating, the median expert group rating and any free text comments in the hope of ratings
9
10 362 reaching a consensus. In this round they will be asked to select the most appropriate instrument for
11
12 363 each outcome in a binary format.

13 364

14
15
16 365 Analysis of Stage 2, Round 3

17
18 366 A final instrument set matched to the core outcome set will be formed based on the consensus. In any
19
20 367 areas where there is no consensus, the study team will adjudicate, taking account of free text
21
22 368 comments.

23
24 369

25
26 370 Data Storage

27
28
29 371 Participants' contact details, including email addresses and telephone numbers, and the answers they
30
31 372 provide, will only be stored by Clinivo for the duration of the study. Clinivo will not share
32
33 373 participants' contact details with any third parties and participants' answers will be stored
34
35 374 anonymously. Data will be encrypted before being stored on Clinivo's server and prior to being
36
37 375 transferred to the University of Glasgow. On completion of the study, Clinivo will destroy all data
38
39 376 after transferring it to the University of Glasgow. The University will securely store the data on
40
41 377 password access computers for a period of ten years following completion of the research project.

42 378

43
44 379

45
46 380

47
48
49 381

50
51 382

52
53 383

54
55
56 384

57
58
59 385
60

1
2
3 386 **Ethics**
4

5
6 387 Ethical approval for this study has been received from the University of Glasgow College of Medical,
7 388 Veterinary and Life Sciences Ethics Committee.
8
9

10 389

11
12 390 **Dissemination**
13

14
15 391 With regard to disseminating the results of our study, we will communicate our results via peer review
16 392 publication, conference presentations, professional societies and also via our institution's social media
17 393 platforms.
18
19

20
21 394 In addition, we will submit a report to our funding body, the Chief Scientist Office of the Scottish
22 395 Government Health Department. We will also share early findings with PHE and Health Scotland. We
23 396 will be in full discussion with both bodies to ensure that our work informs their evaluation plans for
24 397 BWMPs for adults with overweight and obesity.
25
26
27

28 398 Our study is, of course, restricted to the UK. This is due to BWMPs and their settings within health
29 399 services being fairly country-specific. For example, in France and the Netherlands there is no health
30 400 insurance funding of BWMPs and, in the USA, obesity services are tertiary, combining behavioural
31 401 programmes with medication and bariatric surgery. In addition, instruments, such as language and
32 402 health economic models, can be country-specific. Therefore, if used in an international context for
33 403 trials or real world services, our core outcome and definition/instrument set may require further
34 404 adaptation.
35
36
37
38
39

40 405

41 406

42 407

43 408

44 409

45 410

46 411

47 412
48
49
50
51
52
53
54
55
56
57
58
59
60

413 REFERENCES

414

415 (1) National Institute for Health and Care Excellence. Weight Management: lifestyle services for
416 overweight or obese adults (PH53). 2014.

417

418 (2) Scottish Intercollegiate Guidelines Network. Management of Obesity: A national clinical
419 guideline. 2010.

420

421 (3) Loveman E, Frampton GK, Shepherd J, Picot J, Cooper K, Bryant J et al. The clinical
422 effectiveness and cost-effectiveness of long-term weight management schemes for adults: a
423 systematic review. *Health Technol Assess* 2011; 15(2):1-182.

424 (4) Read S, Logue J. Variations in weight management services in Scotland: a national survey of
425 weight management provision. *J Public Health (Oxf)* 2016; 38(3):e325-e335.

426 (5) Public Health England. National mapping of weight management services 2015. 2018.

427

428 (6) Public Health England. Standard Evaluation Framework for weight management
429 interventions. 2009.

430

431 (7) Hartmann-Boyce J, Johns D, Aveyard P, Lewis A, Jebb S, Phillips D et al. Managing overweight
432 and obese adults: The clinical effectiveness of long-term weight management schemes for
433 adults. 2013.

434

435 (8) [https://www.gov.uk/government/publications/adult-weight-management-services-collect-](https://www.gov.uk/government/publications/adult-weight-management-services-collect-and-record-data)
436 [and-record-data](https://www.gov.uk/government/publications/adult-weight-management-services-collect-and-record-data). 2018.

437

438 (9) [https://www.gov.uk/government/publications/adult-weight-management-key-](https://www.gov.uk/government/publications/adult-weight-management-key-performance-indicators)
439 [performance-indicators](https://www.gov.uk/government/publications/adult-weight-management-key-performance-indicators). 2018.

440

441 (10) Coulman KD, Hopkins J, Brookes ST, Chalmers K, Main B, Owen-Smith A et al. A Core Outcome
442 Set for the Benefits and Adverse Events of Bariatric and Metabolic Surgery: The BARIACT
443 Project. *PLoS Med* 2016; 13(11):e1002187.

444

445 (11) Fitch K, Bernstein SJ, Aguilar DM, Burnand B, LaCalle JR, Lazaro P et al. The RAND/UCLA
446 Appropriateness Method User's Manual. 2001. RAND, Santa Monica.

447

448 (12) Park RE, Fink A, Brook RH, Chassin MR, Kahn KL, Merrick NJ et al. Physician ratings of
449 appropriate indications for six medical and surgical procedures. *Am J Public Health* 1986;
76(7):766-772.

450

451

452

453

454

455

456

457

458

- 1
2
3 450 (13) Eldridge SM, Lancaster GA, Campbell MJ, Thabane L, Hopewell S, Coleman CL et al. Defining
4 451 Feasibility and Pilot Studies in Preparation for Randomised Controlled Trials: Development of
5 452 a Conceptual Framework. *PLoS One* 2016; 11(3):e0150205.
- 7 453 (14) Froud R, Eldridge S, Kovacs F, Breen A, Bolton J, Dunn K et al. Reporting outcomes of back pain
8 454 trials: a modified Delphi study. *Eur J Pain* 2011; 15(10):1068-1074.
- 10 455 (15) Petrou S, Rivero-Arias O, Dakin H, Longworth L, Oppe M, Froud R et al. Preferred reporting
11 456 items for studies mapping onto preference-based outcome measures: The MAPS statement.
12 457 *J Med Econ* 2015; 18(11):851-857.
- 15 458 (16) Pincus T, Miles C, Froud R, Underwood M, Carnes D, Taylor SJ. Methodological criteria for the
16 459 assessment of moderators in systematic reviews of randomised controlled trials: a consensus
17 460 study. *BMC Med Res Methodol* 2011; 11:14.
- 19 461 (17) Pinnock H, Barwick M, Carpenter CR, Eldridge S, Grandes G, Griffiths CJ et al. Standards for
20 462 Reporting Implementation Studies (StaRI) Statement. *BMJ* 2017; 356:i6795.
- 23 463 (18) Taylor WJ, Schumacher HR, Jr., Baraf HS, Chapman P, Stamp L, Doherty M et al. A modified
24 464 Delphi exercise to determine the extent of consensus with OMERACT outcome domains for
25 465 studies of acute and chronic gout. *Ann Rheum Dis* 2008; 67(6):888-891.
- 26 466
27 467
28
29
30 468

1
2
3 469 **AUTHORS' CONTRIBUTIONS**
4

5
6 470 RMM and JL drafted the protocol. LJE and SAS critically reviewed the protocol. RMM and JL finalised
7 471 the protocol.
8
9

10 472

11
12 473
13

14
15 474 **FUNDING STATEMENT**
16

17 475 This work was supported by the Chief Scientist Office of the Scottish Government Health Department,
18 476 grant reference number CGA/17/08.
19

20
21 477 SAS was supported by a MRC Strategic Award (MC-PC-13027, MC_UU_12017_14 and SPHSU14).
22

23 478
24

25
26 479
27

28
29 480 **COMPETING INTERESTS STATEMENT**
30

31 481 JL leads a joint working project between University of Glasgow, NHS Greater Glasgow and Clyde, MSD
32 482 and Astra Zeneca. The project also involved an educational grant from Janssen. JL received funding to
33 483 attend a conference from Novo Nordisk.
34

35
36 484 LJE has a part time secondment with PHE.
37
38

39 485
40

41
42 486
43

44
45 487 **FIGURE LEGENDS**
46

47 488 **Figure 1. Schematic outlining the two stage Delphi study.** In order to develop a core outcome set and
48 489 definition/instrument set, Delphi methodology will be used to gain consensus from expert groups.
49 490 Two Delphis (stage 1 and stage 2) will be carried out online over three rounds of questionnaires. The
50 491 stage 1 Delphi will focus on development of a core outcome set. The stage 2 Delphi will focus on
51 492 corresponding definition/instrument selection. PHE, Public Health England; SEF, standard evaluation
52 493 framework; KPI, key performance indicator.
53
54
55
56

57
58 494
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

495

496

For peer review only

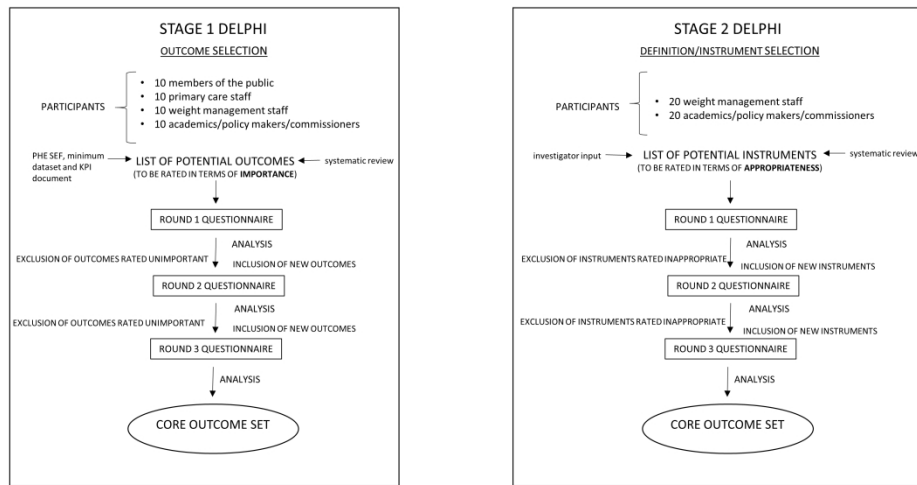


Figure 1. Schematic outlining the two stage Delphi study. In order to develop a core outcome set and definition/instrument set, Delphi methodology will be used to gain consensus from expert groups. Two Delphis (stage 1 and stage 2) will be carried out online over three rounds of questionnaires. The stage 1 Delphi will focus on development of a core outcome set. The stage 2 Delphi will focus on corresponding definition/instrument selection. PHE, Public Health England; SEF, standard evaluation framework; KPI, key performance indicator.

338x190mm (300 x 300 DPI)