Table 1: levels of quality assigned by the GRADE approach to assessing the confidence that can be assigned to the pooled effect estimate from a pairwise meta-analysis. “Current definition” adopted in 2011 series of articles published in Journal of Clinical Epidemiology. “Previous definition” used in 2008 BMJ series of GRADE articles (see reference 5-7 for further information).

Quality level	Current definition	Previous definition
High	We are very confident that the true effect lies close to that of the estimate of the effect	Further research is very unlikely to change our confidence in the estimate of effect
Moderate	We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
Low	Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect	Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
Very low	We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect	Any estimate of effect is very uncertain

Table 2a weight loss programs: Summary results (difference in mean weight loss (kg)) and GRADE assessment of the direct, indirect and NMA analyses. Adapted from [14]. Direct estimates were reported by original authors as being based on Der Simonian and Laird [37]; Network estimates were reported as based on hierarchical Bayesian network meta-regression [38] "accounting for exercise and behavioural support"[14].

Comparison (active vs control)	**Mean difference (95% CI)**			**Quality of Evidence (GRADE)**		
	Direct	**Indirect**	**Network**	Direct	Indirect	*Network*
LEARN vs No diet	3.67 (-3.88, 11.21)	3.63 (0.36, 6.91)	5.16 (2.68, 7.63)	Low	Low	*Low*
Moderate vs No diet	4.84 (2.82, 6.86)	4.69 (1.73, 7.75)	5.70 (4.14, 7.35)	Low	Low	*Moderate*
Low Carb vs No Diet	9.34 (7.31, 11.37)	5.16 (2.25, 8.18)	7.25 (5.33, 9.25)	Low	Moderate	*Moderate*
Low fat vs No diet	5.97 (2.01, 9.92)	6.15 (2.96, 9.40)	7.27 (5.26, 9.34)	Moderate	Moderate	*Moderate*
Moderate vs LEARN	0.21 (-4.64, 5.05)	0.94 (-1.74, 3.66)	0.55 (-1.71, 2.87)	Low	Low	*Low*
Low Carb vs LEARN	1.23 (-1.22, 3.67)	2.48 -0.19, 5.19)	2.10 (-0.20, 4.47)	Low	Low	*Low*
Low fat vs LEARN	4.00 (-0.21, 8.21)	2.64 (-0.02, 5.33)	2.12 (-0.33, 4.59)	Low	Low	*Low*
Low Carb vs Moderate	1.07 (0.16, 1.97)	2.05 (-0.92, 4.96)	1.55 (0.13, 2.95)	Moderate	Low	*Moderate*
Low fat vs Moderate	1.84 (0.96, 2.72)	1.38 (-0.75, 3.51)	1.56 (-0.17, 3.30)	Moderate	Low	*Moderate*
Low fat vs Low CHO	0.33 (-0.86, 1.52)	0.39 (-1.92, 2.70)	0.02 (-1.78, 1.79)	Low	Moderate	*Moderate*

Table 2b Osteoporotic hip fractures GRADE NMA assessment of the direct, indirect and NMA analyses. Estimates are odds ratios (OR), where OR<1 favours active treatment. Adapted from [8]. Only comparisons for which direct data was available are shown here. Original authors report direct estimates as based on random effects models estimated by Comprehensive Meta-analysis version 2 [39]. Network estimates were reported [15] as being based on Bayesian random effects NMA using methods of Lu and Ades [3]

Comparison (active vs control)	Odds ratio (95% CI)			Quality of Evidence (GRADE)		
	Direct	Indirect	Network	Direct	Indirect	Network
Raloxifene *v* placebo	0.84 (0.63 to 1.13)	0.96 (0.53 to 1.78)	0.87 (0.63 to 1.22)	Moderate	Low	*Moderate*
Risedronate *v* placebo	0.17 (0.05 to 0.59)	0.54 (0.36 to 0.75)	0.48 (0.31 to 0.66)	Low	Low	*Low*
Vitamin D *v* placebo	1.25 (0.82 to 1.89)	1.08 (0.61; 1.91)	1.13 (0.94 to 1.34)	Low	Low	*Low*
Vitamin D+calcium *v* placebo	0.83 (0.73 to 0.96)	0.54 (0.29 to 0.94)	0.81 (0.68 to 0.96)	Moderate	Low	*Moderate*
Vit D+calcium *v* teriparatide	2.00 (0.50 to 8.33)	-	1.92 (0.45 to 8.42)	Low	-	*Low*
VitD+calcium *v* denosumab	1.67 (1.02 to 2.70)	-	1.64 (0.97 to 2.87)	Moderate	-	*Moderate*
Alendronate *v* raloxifene	0.49 (0.04 to 5.45)	0.53 (0.30 to 0.90)	0.51 (0.29 to 0.87)	Low	Moderate	*Moderate*
Vit D+calcium *v* raloxifene	0.88 (0.51 to 1.54)	0.96 (0.63 to 1.49)	0.94 (0.66 to 1.31)	Moderate	Low	*Moderate*
ViitD+calcium *v* zoledronate	1.64 (1.16 to 2.17)	-	1.63 (1.16 to 2.30)	High	-	*High*
Vit D+calcium *v* risedronate	1.92 (0.84 to 4.35)	5.88 (1.79 to 25.00)	1.69 (1.27 to 2.54)	Very low	Low	*Low*
VitD+calcium *v* ibandronate	1.72 (0.76 to 3.85)	-	1.69 (0.69 to 3.84)	Low	-	*Low*
Vitamin D *v* alendronate	3.70 (1.20 to 11.11)	2.38 (1.49 to 3.85)	2.54 (1.63 to 4.16)	Moderate	Moderate	*Moderate*
Vit D+calcium *v* alendronate	1.59 (1.03 to 2.44)	2.78 (1.14 to 8.33)	1.82 (1.24 to 2.90)	Moderate	Moderate	*Moderate*
Calcium *v* alendronate	4.55 (0.47 to 50.00)	2.56 (1.54 to 4.35)	2.56 (1.57 to 4.34)	Very low	Moderate	*Moderate*
VitD+calcium *v* vitamin D	1.03 (0.68 to 1.54)	0.65 (0.48 to 0.85)	0.72 (0.57 to 0.91)	Low	Low	*Low*
Calcium *v* calcium+vit D	1.21 (0.89 to 1.66)	3.43 (0.26 to 160.4)	1.40 (1.03 to 1.95)	Low	Very low	*Moderate*

Table 3. Base-case NMA based on the two-stage method, posterior summaries. The highlighted treatment, which would be recommended on the base-case analysis, is the one with the highest ranked mean treatment effect

(a) Branded weight loss programs

Treatment	Pr(Best)	Mean kg difference	SD
No diet	0	(reference)	-
LEARN	0.01	5.56	1.16
Moderate	0	6.09	0.72
Low Carb	0.17	7.49	0.72
Low Fat	**0.82**	**7.88**	**0.76**

(b) Hip fracture treatments in osteoporosis

Treatment	Pr(Best)	Log odds ratio	SD
Placebo	0.00	(reference)	-
Teriparatide	0.32	-0.87	0.72
Denosumab	0.04	-0.69	0.26
Raloxifene	0.00	-0.15	0.13
Zoledronate	0.02	-0.68	0.17
Risedronate	**0.45**	**-1.12**	**0.35**
Ibandronate	0.12	-0.72	0.42
Alendronate	0.05	-0.75	0.21
VitD	0.00	0.04	0.15
VitD+Calcium	0.00	-0.18	0.07
Calcium	0.00	0.02	0.17

Table 4a: Threshold analysis for branded weight loss programs [14] new recommended treatment (original was Low Fat), threshold at which new recommendation is made, and posterior residual mean deviance of the adjusted data NMA model at the threshold adjustment. The GRADE NMA assessment is from the last column in Table 2a. n.c. indicates No Change. n.f. (not found) indicates no threshold was found within + or – 5 kg.

Entries in bold indicate evidence sources in which a plausible bias could change the treatment decision from Low Fat. For example, if the pair-wise evidence Low Carb vs Moderate was subject to a bias of -1kg or more, (ie the unbiased estimate was not the observed 1.07 kg, but 2.07 kg or higher), the treatment recommendation would change to Low Carb.

Treatment B (active)	Treatment A (control)	Estimate (B relative to A)	S.E.	Trials	Recommendation	Bias Threshold, kg	Deviance	GRADE NMA
Learn	No diet	3.67	3.85	2	n.c	n.f	-	Low
Moderate	No diet	4.84	1.03	7	n.c	n.f	-	Moderate
Low carb	No diet	9.34	1.04	1	n.c	n.f	-	Moderate
Low fat	No diet	5.97	2.02	3	n.c	n.f	-	Moderate
Moderate	Learn	0.21	2.47	2	n.c	n.f	-	Low
Low carb	Learn	1.23	1.25	2	Learn	4.5	20.4	Low
Low fat	Learn	4.00	2.15	2	n.c	n.f	-	Low
Low carb	Moderate	1.07	0.46	10	**Low Carb**	**-1.0**	9.9	Moderate
Low fat	Moderate	1.84	0.45	4	**Low Carb**	**+0.9**	12.1	Moderate
Low fat	Low carb	0.33	0.61	4	**Low Carb**	**+0.9**	13.0	Moderate

Table 4b. Threshold analysis drug treatments to prevent osteoporotic hip fractures [8, 15] Original treatment decision based on two-stage analysis was *risedronate*. For each data input we report the new recommended treatment, the threshold at which new recommendation is made and posterior residual mean deviance of the adjusted data NMA model at the threshold adjustment. The GRADE NMA assessment is from the last column in Table 2b. n.c. indicates no change in recommended treatment. n.f. (not found) indicates no threshold was found within + or – 5 on a log odds ratio scale.

For example; consider the pairwise (direct) evidence on placebo vs risedronate, if the pair-wise evidence was subject to a bias of +0.9 on the log odds ratio (LOR) scale or more, (ie the unbiased estimate was not the observed -1.77, but was -0.87 or higher), the treatment recommendation would change to Teriparatide.

Treatment B (active)	**Treatment A (control)**	**LOR (B relative to A)**	**S.E.**	**Trials**	**Recommendation**	**Bias Threshold (LOR)**	**Deviance**	**GRADE NMA**
Raloxifene	Placebo	-0.17	0.24	1	Raloxifene	-1.5	41.0	Moderate
Risedronate	Placebo	-1.77	0.24	2	**Teriparatide**	**+0.9**	13.8	Low
Vit D	Placebo	0.22	0.69	9	Vit D	-3	92.1	Low
VitD+Calcium	Placebo	-0.19	0.04	8	Teriparatide	-1	39.0	Moderate
VitD+Calcium	Teriparatide	0.69	2.30	1	**Teriparatide**	**+0.3**	15.3	Low
VitD+Calcium	Denosumab	0.51	0.99	1	**Denosumab**	**+0.5**	15.3	Moderate
Alendronate	Raloxifene	-0.71	2.52	1	n.c	n.f	-	Moderate
VitD+Calcium	Raloxifene	-0.13	0.60	2	Raloxifene	+4.5	196.0	Moderate
VitD+Calcium	Zoledronate	0.49	0.74	2	**Zoledronate**	**+0.5**	15.3	High
VitD+Calcium	Risedronate	0.65	1.51	3	**Teriperatide**	**-0.4**	16.9	Low
VitD+Calcium	Ibandronate	0.54	1.42	1	**Ibandronate**	**+0.4**	15.3	Low
VitD	Alendronate	1.31	2.36	1	Alendronate	+3.5	58.0	Moderate
VitD+Calcium	Alendronate	0.46	0.88	7	**Alendronate**	**+0.5**	14.0	Moderate
Calcium	Alendronate	1.52	4.10	1	n.c	n.f	-	Moderate
VitD+Calcium	VitD	0.03	0.53	2	VitD	+2.5	115.2	Low
Calcium	VitD+Calcium	0.19	0.54	4	Calcium	-1.5	18.5	Moderate

Table 5. Relationship between distribution of GRADE NMA quality assessments, and the contrasts to which recommendations are sensitive

	High	Moderate	Low	Very Low	Total
Weight-loss					
All contrasts	0	6	4	0	10
Sensitive contrasts	0	3	0	0	3
Osteoporosis					
All contrasts	1	9	6	0	16
Sensitive contrasts	1	2	4	0	7

Figure 1a. Network of comparisons as described by Johnston [14]. Edge thickness is proportional to the number of trials contributing to that pair-wise contrast. Treatment nodes are not weighted.

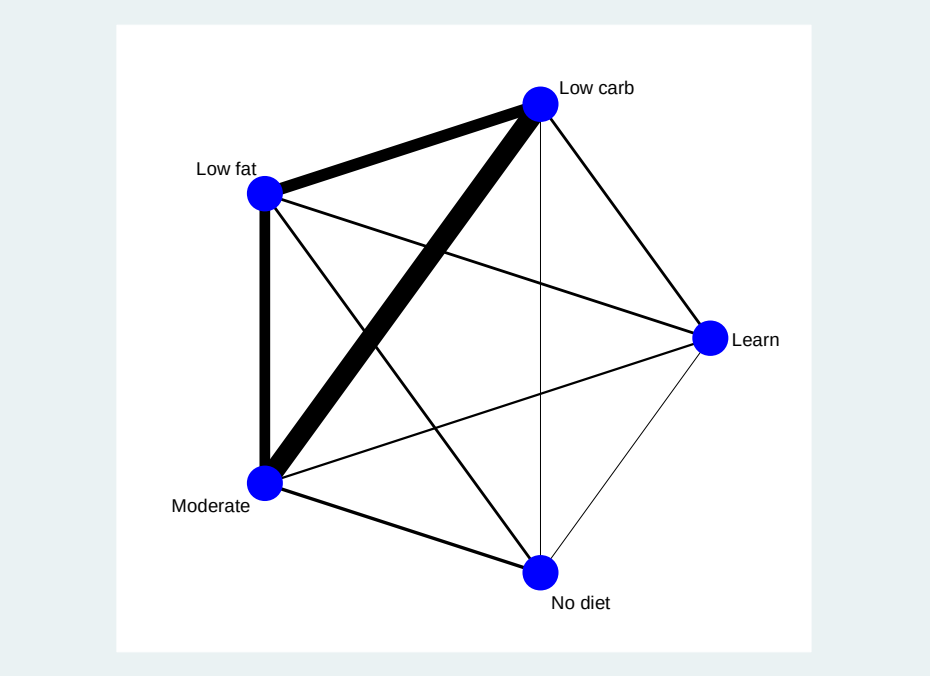

Figure 1b: Osteoporosis network adapted from [8] Edge thickness is proportional to the number of trials contributing to that pair-wise contrast. Treatment nodes proportional to number of participants.