

Appendix A

CHARACTERISTICS OF STUDIES

Characteristics of included studies [ordered by publication year]

Miyazaki et al. 2002

Methods	Prospective cohort study	
Participants	74 patients with medial knee OA	
Joint load exposure	External knee adduction moment during walking assessed at baseline	
Structural knee OA progression outcome	Joint space narrowing grade based on Altman atlas of osteoarthritis. Dichotomous grouping (“progressors”/“non-progressors”) based on at least one grade worsening over the observation period.	
Observation period	72 months	
Quality assessment (risk of bias)		
Quality criteria	Authors’ judgement	Support for judgement
Sufficient description of the groups and the distribution of prognostic factors?	Adequate	Tables adequately presenting relevant information about subjects.
Are the groups assembled at a similar point in their disease progression?	Adequate	Quote: “... the number of patients with each radiographic scale (K/L grade and joint space narrowing scale) was similar in the two groups.”
Is a potential knee joint loading parameter reliably ascertained?	Unclear	There is no mentioning of reliability assessment or reference.
Were the groups comparable on all important confounding factors?	Inadequate	Quote: “However, there were some significant differences at entry between the groups.” Comment:
Was there adequate adjustment for the effects of these confounding variables?	Adequate	The analyses were adequately adjusted (multivariable logistic regression).
Was outcome assessment blind to exposure status?	Adequate	Quote: “All radiographs were evaluated by an experienced reader who was unaware of all other data.”
Was follow-up long enough for the outcomes to occur?	Adequate	Satisfied by default, due to inclusion criteria in the present study.
What proportion of the cohort was followed-up? ^a	Unclear	32 out of a total of 106 (~30%) were lost to follow up.
Were drop-out rates and reasons for drop-out similar across exposed and unexposed groups?	Inadequate	Not described. The exposure for those that were lost to follow up was higher than those who completed the study as “non-progressors”.
Extent of risk of bias in study ^b	Very serious limitation	
Methodological quality	Low quality	>1 criterion being inadequately described.

^a Adequate = follow-up proportion >80%; Unclear 50-80%; Inadequate = <50%

^b Risk of bias is assessed using GRADE’s approach to study limitations: No serious limitation defined as all criteria being adequately described (high methodological quality); Serious limitations defined as one criterion being inadequately described or >1 criterion being unclearly described (moderate methodological quality); Very serious limitation defined as >1 criterion being inadequately described (low methodological quality).

Chang et al. 2007 & Sharma et al. 2003*

Methods	Prospective cohort study	
Participants	57 patients with knee OA (64 patients in abstract) from the MAK cohort (300+ patients).	
Joint load exposure	External knee adduction moment during walking assessed at baseline	
Structural knee OA progression outcome	Joint space narrowing grade based on Altman atlas of osteoarthritis. Dichotomous grouping ("progressing knees"/"non-progressing knees") based on at least one grade worsening over the observation period.	
Observation period	18 months	
Quality assessment (risk of bias)		
Quality criteria	Authors' judgement	Support for judgement
Sufficient description of the groups and the distribution of prognostic factors?	Inadequate	There are no descriptions of the groups, except for the toe-out angle and knee adduction moment.
Are the groups assembled at a similar point in their disease progression?	Inadequate	Not described. Only the distribution of K/L in the full sample is given. The structural disease progression outcome (Altman) is not presented at baseline.
Is a potential knee joint loading parameter reliably ascertained?	Unclear	There is no reliability assessment or reference to previous work.
Were the groups comparable on all important confounding factors?	Inadequate	Not described.
Was there adequate adjustment for the effects of these confounding variables?	Adequate	In the conference abstract there were statistical adjustments for age, gender, BMI, pain, radiographic disease severity (K/L), and varus alignment.
Was outcome assessment blind to exposure status?	Adequate	Quote: "The knee x-ray reader was blinded to the gait analysis data"
Was follow-up long enough for the outcomes to occur?	Adequate	Satisfied by default, due to inclusion criteria in the present study.
What proportion of the cohort was followed-up? ^a	Adequate	This study included 57 at baseline and 56 at follow-up. Note: The underlying MAK cohort is 300+ patients no description of how the gait subgroup was selected
Were drop-out rates and reasons for drop-out similar across exposed and unexposed groups?	Inadequate	The conference abstract (Sharma et al. 2003) included 64 patients in contrast to the 57 patients included in the article (Chang et al. 2007). The authors could not account for the different numbers of subjects.
Extent of risk of bias in study ^b	Very serious limitation	>1 criterion being inadequately described.
Methodological quality	Low quality	

* Conference abstract based on the study.

^a Adequate = follow-up proportion >80%; Unclear 50-80%; Inadequate = <50%

^b Risk of bias is assessed using GRADE's approach to study limitations: No serious limitation defined as all criteria being adequately described (high methodological quality); Serious limitations defined as one criterion being inadequately described or >1 criterion being unclearly described (moderate methodological quality); Very serious limitation defined as >1 criterion being inadequately described (low methodological quality).

Bennell et al. 2011

Methods	Prospective cohort study	
Participants	144 patients with medial knee OA	
Joint load exposure	External knee adduction moment during walking assessed at baseline	
Structural knee OA progression outcome	Progression of 1 or more grades in bone marrow lesions and cartilage defects assessed by semi-quantitative evaluation of MRI scans. Loss in medial tibial cartilage volume assessed from manual segmentation of MRI scans.	
Observation period	12 months	
Quality assessment (risk of bias)		
Quality criteria	Authors' judgement	Support for judgement
Sufficient description of the groups and the distribution of prognostic factors?	Inadequate	There are no descriptions of the groups that progress in BMLs and cartilage defects.
Are the groups assembled at a similar point in their disease progression?	Inadequate	Not described.
Is a potential knee joint loading parameter reliably ascertained?	Unclear	There is no reliability assessment or reference to previous work.
Were the groups comparable on all important confounding factors?	Inadequate	Not described.
Was there adequate adjustment for the effects of these confounding variables?	Adequate	There were statistical adjustments for age, gender, BMI, MRI machine, alignment, baseline values, and treatment group.
Was outcome assessment blind to exposure status?	Inadequate	Not described.
Was follow-up long enough for the outcomes to occur?	Adequate	Satisfied by default, due to inclusion criteria in the present study.
What proportion of the cohort was followed-up? ^a	Unclear	200 enrolled in the underlying study. 144 (72%) had complete data sets.
Were drop-out rates and reasons for drop-out similar across exposed and unexposed groups?	Inadequate	Not described.
Extent of risk of bias in study ^b	Very serious limitation	>1 criterion being inadequately described.
Methodological quality	Low quality	

^a Adequate = follow-up proportion >80%; Unclear 50-80%; Inadequate = <50%

^b Risk of bias is assessed using GRADE's approach to study limitations: No serious limitation defined as all criteria being adequately described (high methodological quality); Serious limitations defined as one criterion being inadequately described or >1 criterion being unclearly described (moderate methodological quality); Very serious limitation defined as >1 criterion being inadequately described (low methodological quality).

Woollard et al. 2011

Methods	Prospective cohort study	
Participants	13 patients with knee OA	
Joint load exposure	External knee adduction moment during walking assessed at baseline	
Structural knee OA progression outcome	Loss of tibiofemoral cartilage volume. Dichotomous grouping (“progressing knees”/“non-progressing knees”) based on a cartilage loss greater than measurement error over the observation period.	
Observation period	12 months	
Quality assessment (risk of bias)		
Quality criteria	Authors’ judgement	Support for judgement
Sufficient description of the groups and the distribution of prognostic factors?	Adequate	Individual patient data is presented according to group structure.
Are the groups assembled at a similar point in their disease progression?	Adequate	Progressors had more severe radiographic disease at baseline (K/L 3 vs. 2), but similar cartilage volumes at baseline (primary outcome).
Is a potential knee joint loading parameter reliably ascertained?	Unclear	There is no reliability assessment or reference to previous work.
Were the groups comparable on all important confounding factors?	Unclear	Progressors had more severe radiographic disease, BMI and lower age; but comparable on gender, WOMAC (all domains) and alignment.
Was there adequate adjustment for the effects of these confounding variables?	Inadequate	The sample was too small for adjustments of the estimates.
Was outcome assessment blind to exposure status?	Inadequate	Not described.
Was follow-up long enough for the outcomes to occur?	Adequate	Satisfied by default, due to inclusion criteria in the present study.
What proportion of the cohort was followed-up? ^a	Adequate	15 invited. 1 lost; 1 excluded.
Were drop-out rates and reasons for drop-out similar across exposed and unexposed groups?	Adequate	Only 1 dropout.
Extent of risk of bias in study ^b	Very serious limitation	>1 criterion being inadequately described.
Methodological quality	Low quality	

^a Adequate = follow-up proportion >80%; Unclear 50-80%; Inadequate = <50%

^b Risk of bias is assessed using GRADE’s approach to study limitations: No serious limitation defined as all criteria being adequately described (high methodological quality); Serious limitations defined as one criterion being inadequately described or >1 criterion being unclearly described (moderate methodological quality); Very serious limitation defined as >1 criterion being inadequately described (low methodological quality).

Henriksen et al. 2013

Methods	Secondary data analysis from a randomized trial	
Participants	157 patients with knee OA and a BMI >30 kg/m ²	
Joint load exposure	Change in peak knee joint compression force during walking over the 4 months prior to the 12 month observation period	
Structural knee OA progression outcome	Boston Leeds Knee Osteoarthritis Score (BLOKS) for tibiofemoral cartilage loss and bone marrow lesions	
Observation period	12 months	
Quality assessment (risk of bias)		
Quality criteria	Authors' judgement	Support for judgement
Sufficient description of the groups and the distribution of prognostic factors?	Adequate	Comment: Demographic, primary outcome and secondary outcome data separated by group are provided in Table 1.
Are the groups assembled at a similar point in their disease progression?	Adequate	Quotation: "There were no other statistically significant differences [including in K&L grade and BLOKS measures of tibiofemoral cartilage and bone marrow lesions] between the Unloaders and Loaders at week 16"
Is a potential knee joint loading parameter reliably ascertained?	Adequate	Quotation: "Based on 30 individuals assessed twice separated by one week done prior to the study, intra-class correlation coefficient for the estimation of peak compression force in our lab is 0.91 (95% CI: 0.83 to 0.96)"
Were the groups comparable on all important confounding factors?	Inadequate	Quotations: "Among the Unloaders the ambulatory knee function was not improved from week 0 to week 16, seen as no statistically significant change in walking speed and a reduction in the knee extensor moment" and "...the ambulatory knee function was improved among the Loaders, seen as increases in walking speed and internal knee extensor moment [from week 0 to week 16]" Comment: These quotations highlight a group difference in the change in walking speed, which may influence the change in knee joint loading which was used to assign group membership.
Was there adequate adjustment for the effects of these confounding variables?	Inadequate	Quotation: "Adjusted analysis: We repeated the analysis further including age, sex, week 16 BMI, and the randomization code in the underlying RCT (Diet, Exercise and Control) as covariates." Comment: Analyses were not adjusted for walking speed or change in walking speed.
Was outcome assessment blind to exposure status?	Adequate	Quotation: "All assessors were blinded to the case status."
Was follow-up long enough for the outcomes to occur?	Adequate	Comment: Satisfied by default, due to inclusion criteria in the present study.
What proportion of the cohort was followed-up? ^a	Adequate	Quotation: "One year after the weight loss (i.e., at week 68), there were 144 subjects (92%) remaining in the biomechanics sub-cohort, i.e., 13 subjects lost to follow-up"
Were drop-out rates and reasons for drop-out similar across exposed and unexposed groups?	Adequate	Quotation: "[participants lost to follow-up] were distributed with 7 (7%) in the Unloader group and 6 (11%) in the Loader group." Comment: Reasons for drop out listed in Figure 2.
Extent of risk of bias in study ^b	Very serious limitation	>1 criterion being inadequately described.
Methodological quality	Low quality	

^a Adequate = follow-up proportion >80%; Unclear 50-80%; Inadequate = <50%

^b Risk of bias is assessed using GRADE's approach to study limitations: No serious limitation defined as all criteria being adequately described (high methodological quality); Serious limitations defined as one criterion being inadequately described or >1 criterion being unclearly described (moderate methodological quality); Very serious limitation defined as >1 criterion being inadequately described (low methodological quality).