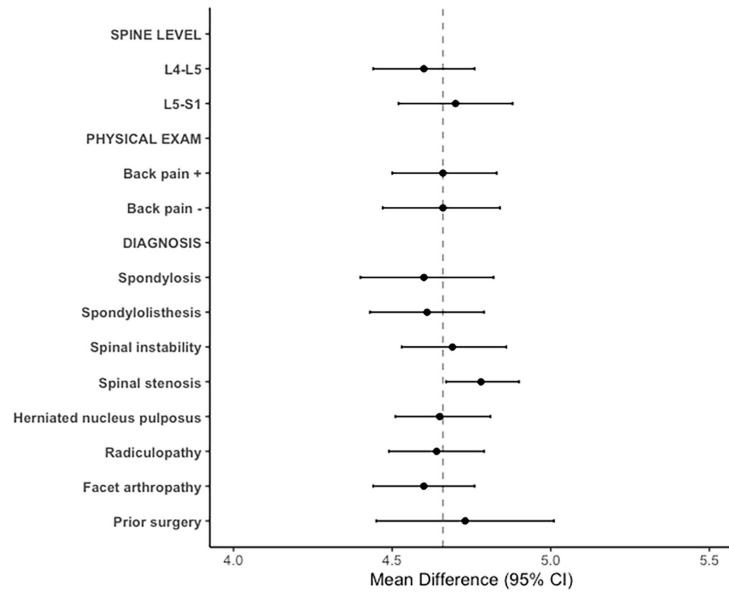
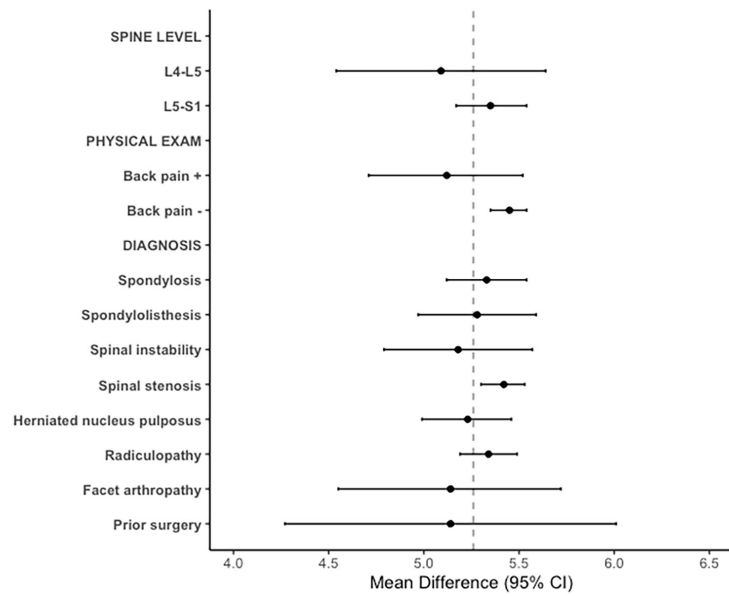


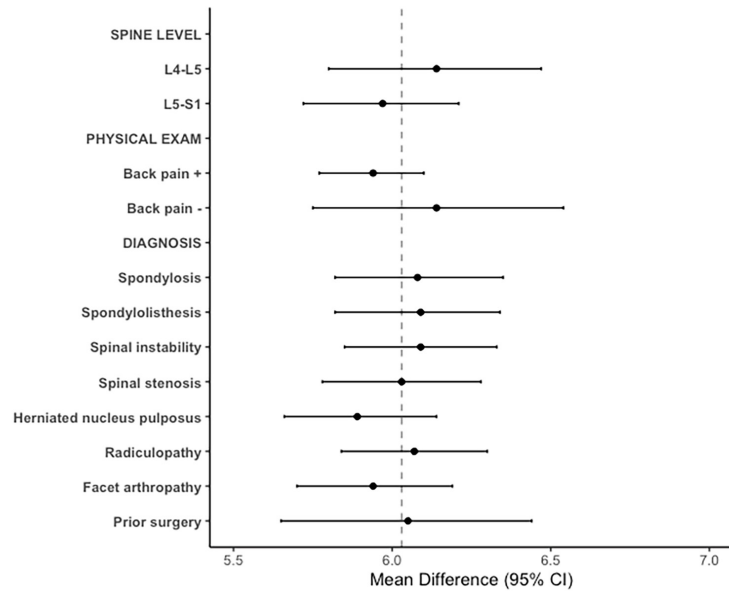
**Figure S1** Change in anterior disc height at 2 years: mean difference between groups by subgroups. CI, confidence interval. Dashed line represents the point estimate of the overall sample. All P values between groups were <0.001.



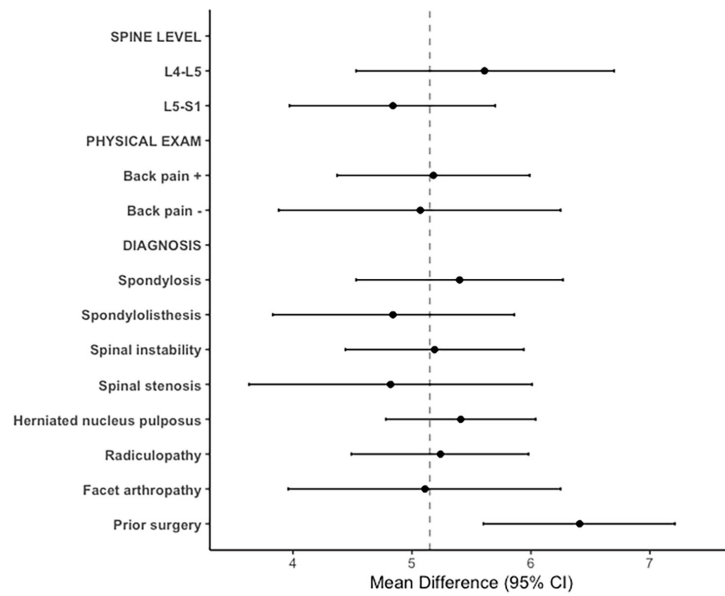
**Figure S2** Change in posterior disc height at 2 years: mean difference between groups by subgroups. CI, confidence interval. Dashed line represents the point estimate of the overall sample. All P values between groups were <0.001.



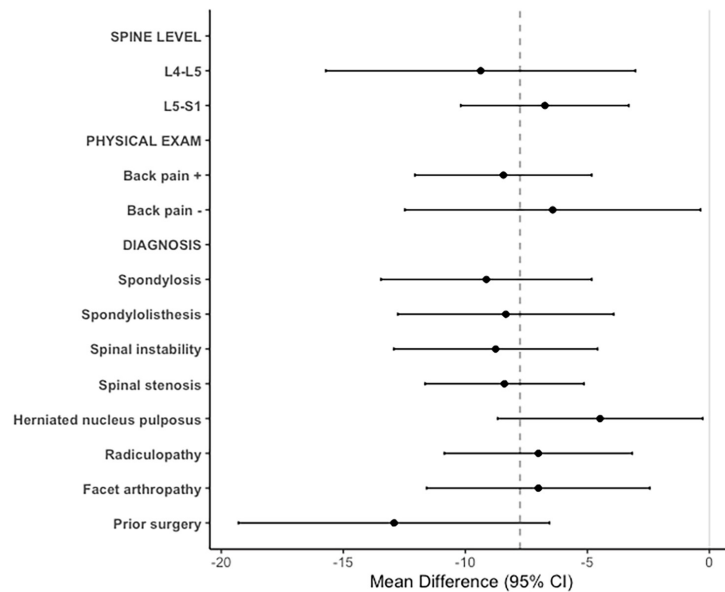
**Figure S3** Change in average disc height at 2 years: mean difference between groups by subgroups. CI, confidence interval. Dashed line represents the point estimate of the overall sample. All P values between groups were <0.001.



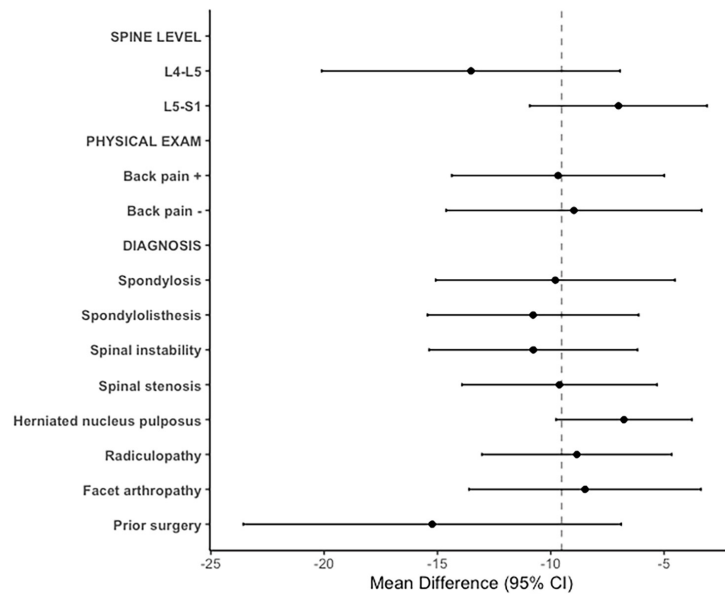
**Figure S4** Change in foraminal height at 2 years: mean difference between groups by subgroups. CI, confidence interval. Dashed line represents the point estimate of the overall sample. All P values between groups were <0.001.



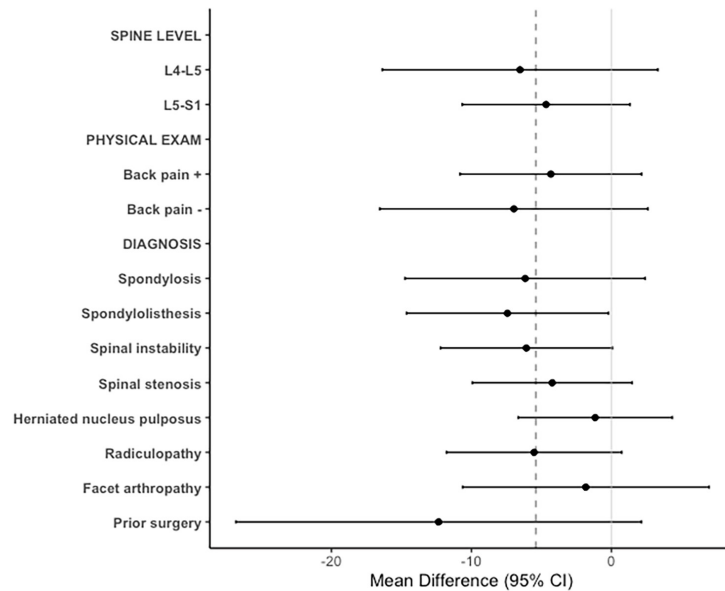
**Figure S5** Change in lordosis at 2 years: mean difference between groups by subgroups. CI, confidence interval. Dashed line represents the point estimate of the overall sample. All P values between groups were <0.001.



**Figure S6** Change in ODI scores at 2 years: mean difference between groups by subgroups. Dashed line represents the point estimate of the overall sample. All P values between groups were <0.001, except for the following subgroups: L4-L5 (P=0.006), back pain -ve (P=0.038), herniated nucleus pulposus (P=0.038), and facet arthropathy (P=0.004).



**Figure S7** Change in VAS back pain at 2 years: mean difference between groups by subgroups. CI, confidence interval. Dashed line represents the point estimate of the overall sample. All P values between groups were  $<0.001$ , except for the following subgroups: back pain -ve ( $P=0.003$ ), facet arthropathy ( $P=0.002$ ), and prior surgery ( $P=0.002$ ).



**Figure S8** Change in VAS leg pain at 2 years: mean difference between groups by subgroups. CI, confidence interval. Dashed line represents the point estimate of the overall sample. All P values between groups were  $>0.05$ , except for the spondylolisthesis subgroup ( $P=0.044$ ).