Statistical Supplement

1. *Analysis of primary and secondary numeric outcomes*. We used longitudinal mixed effects analysis of covariance models to estimate the effects of the randomized treatments on the mean changes from baseline to each follow-up assessment of the primary OSW outcome and numeric outcomes while controlling for the baseline levels of each outcome variable. For each outcome, the mixed model used unstructured covariance matrices to model the serial correlation in responses across the three follow-up visits at 4 weeks, 6 months and 1 year. The model included fixed effects for follow-up time treated as a categorical variable, the baseline level of the outcome, randomized treatment group, and interactions of follow-up time with the randomized treatment and the baseline level. Model parameters were estimated using restricted maximum likelihood estimation.

In sensitivity analyses, the mixed effects analyses of the primary and secondary numeric outcomes were repeated using an as-treated strategy in which 10 of the 110 EPT subjects who did actually receive the intervention were assigned to the usual care group and 5 of the 110 UC subjects who did receive the EPT intervention were assigned to the EPT group.

2. *Multiple imputation* analysis of primary and secondary outcomes. The results of the mixed effects analyses assume that missing data follow a missing data follow a missing at random (MAR) pattern in which the probability of missingness may depend on other observed outcome values in the model, but are not related to the unobserved values of missing responses themselves. To assess the robustness of our results to bias from missing data, we repeated the mixed effects analyses of the primary and secondary numeric outcomes after multiply imputing missing data using fully sequential imputation under an imputation model that incorporated additional auxiliary variables that were selected using subject matter considerations as potential predictors of the outcomes or the probability of missingness. The multiple imputation models included all variables in the analytic model plus the baseline assessments of gender, age, baseline anxiety, depression, concurrent neck or upper back pain, BMI, length in days of current episode of LBP, and the pain catastrophizing scale, as well as indicator variables for surgery or injections prior to the 4 week and 6 month assessments, and nonmissing longitudinal assessments of the OSW, FABQ physical activity and work subscales, EQ-5D quality of life index, EQ-5D self-rating, and the ratings of low back pain and leg pain in the preceding 24 hours. A total of 25 imputed data sets were created. Results were pooled across these 25 data sets using Rubin's formulae to adjust standard errors to account for variation in results between the imputed data sets in order to account for uncertainty resulting from imputation of missing data.

The sensitivity analyses based on multiple imputation retain a MAR assumption, but the assumption is relaxed to allow dependence of missing outcomes and probability of missingness of the auxiliary variables in the imputation model.

3. *Analyses of utilization outcomes and patient-reported success*. Our main analyses of each of the utilization outcomes – surgery, lumbar epidural injection, advanced imaging and emergency department visits - were performed by calculating relative risk statistics and 95% confidence intervals.

We used relative risk rather than Cox regression in our main analyses of each utilization outcome because these outcomes generally occurred early in follow-up and we did not view the timing of the outcome events as highly clinically relevant. In order to assess the dependence of

the results to early dropout and missing assessments, we repeated our analyses of each of the utilization outcomes using Cox regression analysis to relate the outcome to the randomized treatment assignment, with follow-up censored at the final monthly utilization assessment. The Cox regression analyses were performed using a counting process style input in which intermittently missing monthly assessments were excluded from the risk set.

We applied separate analyses of relative risk to relate patient-reported success at the 4 week, 6 month and 1 year assessments to the randomized treatment assignment.

4. *Analyses of missed work days*. We used a generalized estimating equations (GEE) analysis with robust standard errors under a negative binomial outcome model with logarithmic link function to compare the number of missed work days due to LBP per month over the 1 year follow-up period between the randomized groups. The GEE analysis was performed using a compound symmetry working covariance matrix to account for serial correlation over the monthly assessments during the follow-up period. The model included randomized treatment assignment and follow-up month coded as a categorical factor as predictor variables.

A separate GEE analysis used a Bernoulli outcome model with logarithmic link to compare the proportion of patients missing at least one work day in a given month between the randomized treatment groups. This analysis also used a compound symmetry working covariance model and robust standard errors for statistical inference.