The authors propose to use listwise deletion to address missingness of key covariates, including the case worker prognosis variable. Another strategy commonly used in the matching literature would be to separate out subjects missing the caseworker prognosis variable as a stratum unto themselves. The primary outcome analysis might be restricted to subjects who were not missing this covariate, while study of the no-prognosis stratum could be considered as a supplementary analysis, one with a greater potential for unmeasured confounding.

Reviewer 1

The stratification strategy for waiting time seems arbitrary. Were the cutoffs based on quantiles? (Doesn't appear to be based upon sample sizes in table 4) If not, how were they chosen?

Are subjects aware of their prognosis? If so, assumption 3i seems very strong to me. A longer prognosis (whether or not the prognosis is appropriate) might be abused by a subject who could really return to work earlier (e.g. had a speedier then average recovery).

Algorithm steps 2 and 3 - The discarding strategy in these two steps are at odds. Step 3 proceeds to restrict the number of discarded treatments, but step 2 discards at the arbitrary cutoff of needing at least 5 controls. Why not be more flexible with that threshold?

Reviewer 2

Thank you for the opportunity to review this manuscript. Overall, it is a very thoughtfully prepared study protocol. I have a few comments that should be considered in the revision process.

1. Assignment to the treatment group. There is limited information provided concerning how individuals are assigned to the JAM tx group, and no information for the post-match activities of the control group. A richer description of both would allow for a better assessment of the appropriateness of the proposed research design.

2. Rationale for conceptualizing JAM alone as the treatment. The authors describe two types of interventions – work preparatory and work oriented – that are delivered following a JAM if the individual is determined to be ready for services from PES. Assuming that not all individuals who attend a JAM are either (a) determined ready for services or (b) receive services even when deemed eligible, wouldn't receipt of these services constitute a distinct form of treatment? Are these services accessible to individuals from the control group? The answer to the latter question is unclear in the current manuscript.

3. Prognosis variable as proxy. The authors have elected to use caseworker prognosis as a proxy for unobserved confounders. I'm skeptical of the use of caseworker prognosis for this purpose given the possibility for variability across caseworkers in how this determination is made. What information does the caseworker collect that allows for an accurate prognosis that differs in some significant way from estimates based on the proportion of individuals who passed 30, 90 and 180 days of sick leave for a given ICD10 disease? Are you confident that caseworkers are influenced by the same factors and that these factors are accorded equal weight in determining the worker's prognosis? The authors rightly state in the discussion that

the proxy property is not empirically testable but it should be theoretically grounded.

4. Missing data. If the authors elect to continue using the prognosis variable then reporting out the chi-square value for Little's MCAR test related to prognosis would help to instill confidence that the data were in fact missing completely at random, a major issue given the volume of observations removed for this reason.