Supplement 4. Criteria Used in Risk of Bias Assessment

I. Guidance on Assessing Risk of Bias for Randomized Controlled Trials
General instructions: (1) Rate each risk of bias item listed below as Low risk/ High risk/ Unclear risk (refer to Cochrane guidance to inform judgements). Add comments to justify ratings. (2) After considering each quality item, give the study an overall rating of “Low risk,” “Moderate risk,” or “High risk” (see below).

Rating of individual items
* Indicates items contained in Cochrane Risk of Bias Tool.

1. Selection bias:
   a. *Randomization adequate (Adequate methods include random number table, computer-generated randomization, minimization without a random element.) Low risk/ High risk/ Unclear risk
   b. *Allocation concealment (Adequate methods include pharmacy-controlled randomization, numbered sealed envelopes, central allocation.) Low risk/ High risk/ Unclear risk
   c. Baseline characteristics (Consider whether there were systematic differences observed in baseline characteristics and prognostic factors between groups, and if important differences were observed, if the analyses controlled for these differences.) Low risk/ High risk/ Unclear risk

2. Performance bias:
   a. *Concurrent interventions or unintended exposures (Consider concurrent intervention or an unintended exposure (e.g., crossovers; contamination – some control group gets the intervention) that might bias results) Low risk/ High risk/ Unclear risk
   b. Protocol variation (Consider whether variation from the protocol compromised the conclusions of the study.) Low risk/ High risk/ Unclear risk

3. Detection bias:
   a. *Subjects blinded (Consider measures used to blind subjects to treatment assignment and any data presented on effectiveness of these measures.) Low risk/ High risk/ Unclear risk
   b. *Outcome assessors blinded, hard outcomes (Outcome assessors blind to treatment assignment for “hard outcomes” such as mortality.) Low risk/ High risk/ Unclear risk
   c. *Outcome assessors blinded, soft outcomes (Outcome assessors blind to treatment assignment for “soft outcomes” such as symptoms.) Low risk/ High risk/ Unclear risk
   d. Measurement bias (Reliability and validity of measures used.) Low risk/ High risk/ Unclear risk
4. **Attrition bias:**
   a. *Incomplete outcome data* (Consider whether incomplete outcome data were adequately addressed, including systematic differences in attrition between groups [differential attrition]; overall loss to followup [overall attrition]; and whether an “intention-to-treat” [ITT; all eligible patients that were randomized are included in analysis] analysis was performed.) (Note: mixed models and survival analyses are, in general, ITT.) **Low risk/High risk/Unclear risk**

5. **Reporting bias:**
   a. *Selective outcomes reporting* (Consider whether there is any suggestion of selective outcome reporting; e.g., systematic differences between planned and reported findings.) **Low risk/High risk/Unclear risk**

**Overall study rating**
Please assign each study an overall quality rating of “Low risk,” “High risk,” or “Unclear risk” based on the following definitions:

A **“Low risk”** study has the least bias, and results are considered valid. A low risk study uses a valid approach to allocate patients to alternative treatments; has a low dropout rate; and uses appropriate means to prevent bias, measure outcomes, and analyze and report results. [Items 1a and 1c; 2a; 3b and 3c; and 4a are all rated low risk]

A **“Moderate risk”** study is susceptible to some bias but probably not enough to invalidate the results. The study may be missing information, making it difficult to assess limitations and potential problems (unclear risk). As the moderate risk category is broad, studies with this rating vary in their strengths and weaknesses. [Most, but not all of the following items are rated low risk: Items 1a and 1c; 2a; 3b and 3c; and 4a]

A **“High risk”** rating indicates substantial bias that may invalidate the results. These studies have serious errors in design, analysis, or reporting; have large amounts of missing information; or have discrepancies in reporting. The results of a high risk study are at least as likely to reflect flaws in the study design as to indicate true differences between the compared interventions. [At least one-half of the individual quality items are rated high risk or unclear risk]

**Conflict of interest (recorded but not used as part of Risk of Bias Assessment)**
Was there the absence of potential important conflict of interest? The focus here is financial conflict of interest. If no financial conflict of interest (e.g., if funded by government or foundation and authors do not have financial relationships with drug/device manufacturer), then answer “Yes.” **Yes/No/Unclear**
II. Guidance on Assessing Risk of Bias for Nonrandomized Studies
This tool is intended to evaluate the quality of nonrandomized studies that assessed the outcomes of nurse-managed protocol interventions. Use this risk of bias tool for the following study designs: nonrandomized controlled trial, cohort studies, interrupted time series.

Instructions for use:
1. Items are organized by risk of bias domains (selection, performance, attrition, detection and reporting bias). Rate each question using the response categories listed. Focus on study design and conduct, not quality of reporting.

2. The first question, basic study design, is not used in the overall ratings but is collected for descriptive purposes.

3. After answering each item, rate the study overall as “low risk of bias,” “moderate risk of bias,” or “high risk of bias” based on the following definitions. This overall rating is specific to the basic study design used. For example, if the basic study design was a cohort study, then the risk of bias rating would be interpreted as “For a cohort study, the risk of bias is ______.”

A “Low Risk of Bias” study has the least bias, and results are considered valid. A good study has a clear description of the population, setting, interventions, and comparison groups; uses recruitment and eligibility criteria that minimizes selection bias; has a low attrition rate; and uses appropriate means to prevent bias, measure outcomes, and analyze and report results. These studies will meet the majority of items in each domain.

A “Moderate Risk of Bias” study is susceptible to some bias but probably not enough to invalidate the results. The study may be missing information, making it difficult to assess limitations and potential problems. As the fair-quality category is broad, studies with this rating vary in their strengths and weaknesses. The results of some fair-quality studies are possibly valid, while others are probably valid. These studies will meet the majority of items in most but not all domains.

A “High Risk of Bias” rating indicates substantial bias that may invalidate the results. These studies have serious errors in design, analysis, or reporting; have large amounts of missing information; or have discrepancies in reporting. The results of a poor-quality study are at least as likely to reflect flaws in the study design as to indicate true differences between the compared interventions.
1. Basic Design
Is the study design prospective, retrospective, or mixed? [Abstractor: Prospective design requires that the investigator plans a study before any data are collected. Mixed design includes case-control or cohort studies in which one group is studied prospectively and the other retrospectively.]

- Prospective
- Mixed
- Retrospective
- Cannot determine

2. Selection Bias
2.1 Inclusion/exclusion criteria
Did the study apply inclusion/exclusion criteria uniformly to all comparison groups?
- Yes (low risk of bias)
- No (high risk of bias)
- Cannot determine (unclear risk of bias)
- NA: study does not include comparison groups

2.2 Recruitment:
Did the strategy for recruiting participants into the study differ across study groups?
- Yes (high risk of bias)
- No (low risk of bias)
- Cannot determine (unclear risk of bias)
- NA (retrospective study design)

2.3 Baseline characteristics similar or appropriate adjusted analysis
Are key characteristics of study participants similar between intervention and control groups? [Patients Age, Race, Gender, Illness severity] If not similar, did the analyses appropriately adjust for important differences?
- Yes (similar or appropriate adjusted analysis; low risk of bias)
Partially (only some characteristics described or some characteristics not clearly described; analysis adjust for some)

No (important baseline differences, unadjusted analysis; high risk of bias)

2.4 Comparison Group

Is the selection of the comparison group appropriate? [Patients exposed to usual care or another quality improvement strategy is appropriate; if comparison group determined at the physician or practice level, the comparison groups should be drawn from the same system.)

Yes (low risk of bias)

No (high risk of bias)

Cannot determine, no description of the derivation of the comparison cohort (unclear risk of bias)

NA (study does not include a comparison cohort - case series, one-arm study)

2.5 Balance prognostic variables between groups through design or analysis approaches.

Any attempt to balance the allocation between the groups? [For example, through stratification, matching, propensity scores]

Yes (low risk of bias)

No (high risk of bias)

Cannot determine (unclear risk of bias)

3. Performance Bias

3.1 Intervention implementation

Did variation from the study protocol compromise the conclusions of the study [Similar to a psychologist following a manualized procedure to deliver psychotherapy, the nurse-managed protocol intervention should be implemented as planned]?

Unclear (no data reported on fidelity to protocol; unclear risk of bias)
Low fidelity (few components of protocol implemented; high risk of bias)

High fidelity (all key components of protocol were implemented; low risk of bias)

3.2 Concurrent/concomitant interventions

Did researchers rule out any impact from a concurrent intervention, such as greater access to other specialty interventions or medications (e.g., through multivariate analysis, stratification, or subgroup analysis)?

Yes (low risk of bias)

No or Partially (only some concurrent interventions eliminated; high risk of bias)

Not described (unclear risk of bias)

4. Attrition Bias

4.1 Equality of length of followup for participants

In cohort studies, is the length of followup different between the groups? [Abstractor: Where followup was the same for all study patients the answer is no. If different lengths of followup were adjusted by statistical techniques, for example, survival analysis, the answer is no. Studies where differences in followup are ignored should be answered yes.]

Yes (High risk of bias)

No (Low risk of bias)

Cannot determine (Unclear risk of bias)

4.2 Completeness of followup

Was there a high rate of differential or overall attrition? [Attrition is measured in relation to the time between baseline (allocation in some instances) and outcome measurement. Standard for overall attrition is <20 percent for <1 year f/u and <30 percent for longer term ≥ 1 year. Standard for differential attrition is ≥ 10% absolute difference.]

Yes (high risk of bias)

No (low risk of bias)

Cannot determine (unclear risk of bias)
4.3 Attrition affecting Participant Composition
Did attrition result in a difference in group characteristics between baseline and followup?

- Yes (high risk of bias) ☐
- No (low risk of bias) ☐
- Cannot determine (unclear risk of bias) ☐

4.5 Intention-to-treat analysis
Is the analysis conducted on an intention-to-treat (ITT) basis, that is, the intervention allocation status rather than the actual intervention received? [Abstractor: evaluate whether the analysis takes into account loss to followup]

- Yes (low risk of bias) ☐
- No (high risk of bias) ☐
- Cannot determine (unclear risk of bias) ☐
- Not applicable (retrospective study) ☐

5. Detection Bias

5.1 Blind outcomes assessment
Were the outcome assessors blinded to the intervention or exposure status of participants?

- Yes (low risk of bias) ☐
- No or not stated and outcome could be influence by knowledge of exposure status (high risk of bias) ☐
- NA (not an intervention study) ☐

5.2 Source of information re interventions/exposure
Are interventions/exposures assessed using valid and reliable measures, implemented consistently across all study participants?

- Yes (low risk of bias) ☐
- No (high risk of bias) ☐
- Cannot determine, measurement approach not reported (unclear risk of bias) ☐
5.3 **Source of information re outcomes**

a. Are **primary outcomes** (e.g., biophysical measures, performance metrics, symptom/functional status measures) assessed using valid and reliable measures and implemented consistently across all study participants?

- Yes (low risk of bias) □
- No (high risk of bias) □
- Cannot determine, measurement approach not reported (unclear risk of bias) □

b. Are confounding variables assessed using valid and reliable measures, implemented consistently across all study participants? *[Major potential confounders include: age, gender, race, disease severity, overall burden of disease.]*

- Yes (low risk of bias) □
- No (high risk of bias) □
- Cannot determine, measurement approach not reported (unclear risk of bias) □

6. **Reporting Bias**

Are the potential outcomes pre-specified by the researchers? Are all pre-specified outcomes reported? *[Abstractor needs to identify all pre-specified, primary outcomes that should be reported in the study.]*

- Yes (low risk of bias) □
- No (at least 1 pre-specified outcome not reported; high risk of bias) □
- Primary outcomes not pre-specified (unclear risk of bias) □