**Appendix I**

 **Quality of observational studies**

General instructions: Grade each criterion as “Yes,” “No,” “Partially,” or “Can’t tell.” Factors to consider when making an assessment are listed under each criterion. Note that some criteria will only apply to specify types of study. For example, power calculations are relevant for studies aiming to compare suicide risk between two groups, or studies that look at correlates of suicidality/self-harm within an UHR sample. However, power calculations are not relevant in an uncontrolled study of a single UHR sample where suicide related data is only described (rather than featuring in any inferential statistics). Where a criterion only applies to a specific design, it is in italics.

UHR = Ultra High Risk of psychosis

1. **Unbiased selection of the cohort?**

Factors that help reduce selection bias:

* + Inclusion/exclusion criteria
		- Clearly described
		- Criteria for achieving UHR status clearly outlined or previous literature outlining these criteria are referred to.
	+ Recruitment strategy
		- Clearly described
		- Sample is representative of the population of interest: Help-seeking UHR individuals. Samples who are not help-seeking may not be representative of this group.
1. ***Selection minimizes baseline differences in prognostic factors (For controlled studies only)?***

Factors to consider:

* + Was selection of the comparison group appropriate? Consider whether these two sources are likely to differ on factors related to the outcome (besides UHR status). Note that in instances of UHR versus non-clinical controls, differences in clinical characteristics would be expected, but matching on key demographics (age, gender, ethnicity, education, etc.) would still be required to minimize bias.
	+ Did the study investigators do other things to ensure that exposed/unexposed groups were comparable, e.g., by using stratification or propensity scores?
1. ***Sample size calculated (for controlled studies and where studies test for predictors/correlates of suicidality/self-harm)?***

Factors to consider:

* + Did the authors report conducting a power analysis or describe some other basis for determining the adequacy of study group sizes for the primary outcome(s) of interest to us?
	+ Did the eventual sample size deviate by < 10% of the sample size suggested by the power calculation?
1. **Adequate description of the cohort?**

Consider whether the cohort is well-characterized in terms of baseline demographics?

* + Consider key demographic information such as age, gender and ethnicity.
	+ Information regarding education or socio-economic characteristics is also important.
1. **Validated method for ascertaining UHR status?**

Factors to consider:

* + Was the method used to ascertain exposure clearly described? (Details should be sufficient to permit replication in new studies)
	+ Was a valid and reliable measure used to ascertain exposure? (self-report measures tend to have lower reliability and validity than clinical interview). Gold standard tools include the Comprehensive Assessment of the At-Risk Mental State (CAARMS) and the Structured Interview of Prodromal Syndromes (SIPS).
1. **Validated method for ascertaining suicidality or self-harm?**

Factors to consider:

* + Were primary outcomes assessed using valid and reliable measures? Note that measures that consist of single items of scales taken from larger measures are likely to lack content validity and reliability.
	+ Were these measures implemented consistently across all study participants?
1. **Outcome assessment blind to exposure ?**
	* Were the study investigators who assessed outcomes blind to the UHR status of participants? (Note that even in single-arm studies so degree of blinding is possible, for example using external interviewers with no knowledge of participants clinical status).
2. ***Adequate follow-up period (longitudinal studies only)?***

Factors to consider:

* + Minimum adequate follow-up period is 1-year for suicide attempts. A shorter follow-up period may be appropriate where suicidal ideation is the outcome. A longer period will be required where completed suicide is the outcome.
	+ A justification of the follow-up period length is preferable.
	+ Follow-up period should be the same for all groups
		- OK if differences in follow-up time were adjusted for using statistical techniques, e.g., survival analysis.
1. **Missing data**

Factors to consider:

* + Did missing data from any group exceed 20%?
	+ In longitudinal studies consider attrition over time as a form of missing data. Note that the criteria of < 20% missing data may be unrealistic over longer follow-up periods.
	+ If missing data is present and substantial, were steps taken to minimize bias (e.g., sensitivity analysis or imputation).
1. ***Analysis controls for confounding (controlled studies and where studies test for predictors/correlates of suicidality or self-harm)?***

Factors to consider for controlled studies:

* + Does the study identify and control for important confounding variables and effect modifiers? Confounding variables are risk factors that are correlated with UHR status and outcome and may therefore bias the estimation of the effect of UHR status on outcome if unmeasured. These may include demographic and clinical variables (e.g., co-morbidity).

Factors to consider for studies looking at predictors of suicide risk within UHR groups:

* Did the study control for likely demographic and clinical confounders? For example, using multiple regression to adjust for demographic or clinical factors likely to be correlated with predictor and outcome?
1. ***Analytic methods appropriate (Controlled studies and where studies test for predictors/correlates of suicidality or self-harm)?***

Factors to consider:

* + Was the kind of analysis done appropriate for the kind of outcome data (categorical, continuous, etc.)?
	+ Was the number of variables used in the analysis appropriate for the sample size? (The statistical techniques used must be appropriate to the data and take into account issues such as controlling for small sample size, clustering, rare outcomes, multiple comparison, and number of covariates for a given sample size)