Critical Appraisal Process for Assessment of Public Health Measures for COVID-19 Cohort Studies

We appraised the quality of the individual studies using an adapted version of ROBINS-I. This tool classifies the Risk of Bias of a study as **Low, Moderate, Serious, Critical, or No Information**. Low Risk of Bias indicates High Quality, and Critical Risk of Bias indicates Very Low (insufficient) Quality. ROBINS-I appraises 7 bias domains and judges each study against an ideal reference randomized controlled trial. To improve the utility of ROBINS-I for assessing studies reporting on public health measures for COVID-19, we have focused on study characteristics that introduce bias as reported in the COVID-19 literature. Questions associated with each ROBINS-I domain and associated judgements were decided by consensus among the authors of the Living Evidence Syntheses (Table). An overall judgement of "serious" or "critical" is given when the study is judged to be at critical risk of bias in at least one domain. Three or more serious risk of bias domains is given an overall risk of bias of critical.

VE Study Characteristics that may introduce bias	Description
ROBINS-I: Bias due to confounding	**Did the study adjust for other COVID protective interventions (including vaccination)?** *Critical* = multiple co-interventions with no controlling or adjustment *Serious* = one co-intervention not controlled for *Moderate* = all known important interventions controlled for **Did the study adjust for calendar time (implications for circulating variant, season), demographics, and other relevant factors?** *Critical* = no adjustment *Serious* = at least one known important domain not measured or controlled for *Moderate* = all known important confounding domains measured **Were participants free of confirmed COVID infection at the start of the study?** *Critical* = unclear or high likelihood participants had COVID at start of study *Serious* = COVID status of intervention group known but unclear for control group OR COVID status of both groups known by self-report only *Low* = negative COVID status of both groups known at study start (laboratory test confirmed)
ROBINS-I: Bias in selection of participants into study	**Were both study groups recruited from the same population during the same time period?** *Critical* = same or different country/province/state measured at a different time prior to pandemic *Serious* = same or different country/province/state measured at a different time during pandemic *Moderate* = same country/province/state measured at same time **Were the COVID protective interventions implemented prior to period of data collection?** (prevalent users) *Critical* = not addressed and highly likelihood of prevalent users *Moderate* = prevalent users likely but appropriately controlled for *Low* = start of data collection at same time as implementation with no prevalent users **Were the study groups balanced with respect to participant adherence (based on internal and external factors unrelated to COVID)?** For example, people who are less likely to adhere to public health measures anyway may be more likely to be exposed to COVID and require quarantine & isolation but then are

	less likely to adhere. Similar for e.g. people who work are essential workers without paid time off. *Critical* = not addressed and highly likelihood of difference in adherence *Moderate* = difference in adherence likely but appropriately controlled for *Low* = adherence confirmed to be same in both groups at start of study
ROBINS-I: Bias in classification of interventions	**Was the method for confirming the intervention clearly defined and applied consistently across study samples (e.g., districts within a country)?** *Critical* = not addressed *Serious* = intervention status not well defined or applied inconsistently *Moderate* = well defined but some aspects of assignment of intervention status determined retrospectively *Low* = well defined and solely based on information collected at time of intervention **In periods of co-occurring interventions, do the authors clearly classify each individual intervention?** *Critical* = not addressed and co-interventions present *Serious* = co-intervention classification not well defined or applied inconsistently *Moderate* = co-intervention classification well defined but some aspects of assignment of status determined retrospectively *Low* = all co-interventions well defined and solely based on information collected at time of intervention **Does classification into intervention/control group depend on self-report in a way that might introduce bias?** For example, where negative consequences of providing truthful responses may lead to negative consequences e.g. self-reporting COVID symptoms would trigger 14 day quarantine and loss of income *Critical* = not addressed and reliant on self-report *Moderate* = reliant on self-report but appropriately controlled for/analyzed separately *Low* = not reliant on self-report **For household transmission studies, was it clear that exposure to the index case was the most likely the only exposure to COVID for household or close contacts?** *Critical* = not addressed *Serious* = high risk occupational and social exposures likely and not accounted for *Moderate* = all participants isolated to same house or hospital from time of index case identification *Low* = all participants isolated to same house or hospital prior to index case identification
ROBINS-I: Bias due to deviations from intended intervention	**Did the authors assess adherence to the protective behaviours/interventions after intervention implementation?** *Critical* = not addressed *Serious* = reliant on self-report of adherence without verification or adjustment *Moderate* = adherence verified in at least a subset of each study group or appropriately adjusted for *Low* = adherence verified in all study participants
ROBINS-I: Bias due to missing data	**Was outcome data at the end of the study period available for all or nearly all participants?** *Critical* = critical differences in missing data between groups *Moderate* = missing data did not differ between groups or was accounted for by appropriate statistical methods

	Low = no missing data **Were participants excluded due to missing data?** *Critical* = participants excluded based on data missing unevenly across groups *Moderate* = participants excluded due to missing data, but rationale was appropriate and applied the same across all groups *Low* = no exclusions due to missing data
ROBINS-I: Bias in measurement of outcomes	**Was the outcome of COVID confirmed by laboratory testing?** *Critical* = not reported *Serious* = only sample or subset of population had PCR *Moderate* = most participants had PCR *Low* = all participants had PCR **If the outcomes were derived from databases, were the databases constructed specifically for the collection of COVID data?** *Critical* = no or unclear *Serious* = database for non-COVID purpose without individual level data *Moderate* = database for non-COVID purpose with individual level data (e.g. health records, employee records) *Low* = national/state/province level surveillance database or specifically for COVID **Were appropriate tools/methods with validated/justified cut-points used to determine outcomes of interest (other than COVID infection/transmission which is covered under laboratory testing)?** *Critical* = not reported *Serious* = outcomes solely dependent on self-report without a validated measure *Moderate* = objective measure applied but validation uncertain *Low* = objective validated measure used consistently across all groups **If the outcome was self-reported, did the authors attempt to control for social desirability?** *Critical* = not reported and outcome likely to be influenced by social desirability *Moderate* = attempt made to control for social desirability *Low* = outcome not influenced by social desirability **Was the frequency of testing for the outcome different between the study groups?** *Critical* = routinely done more frequently in one group more than the other *Moderate* = some differences but rationale appropriate *Low* = no difference in frequency of testing between groups **If outcome was observed, was there more than one assessor and if so, was interrater agreement reported?** *Critical* = not reported *Serious* = reported with low agreement *Moderate* = reported with moderate agreement *Low* = reported with excellent agreement

Citation: Linkins LA. Critical appraisal process for assessment of public health measures for COVID-19 cohort studies. Hamilton, Canada: Health Information Research Unit, 22 March 2023.