Box 1. Tool for assessing susceptibility to bias in observational studies

Domain	Tool item must address	Met domain criteria?#				
		No	**Partially**	**Yes**	**?**	**NA**
	MAJOR CRITERIA					
***Methods for selecting study participants**	*The selection procedures (inclusion and exclusion criteria, sources and selection methods) are appropriate to represent the universe of interest?* Ideally sampling should be probabilistic (all participants have a known probability of being selected). Did the study use an appropriate sampling method to avoid selection bias. The risk of bias is high in convenience samples or if the technique of sampling or selection of controls is not clear.					
	Is the sample a representative spectrum of the population of interest? Take into account how the study sample is representative of the population of interest (general population, admitted patients, outpatients in a health center, age group, schooled). The risk of bias increases in studies conducted in special populations. In case-control studies, cases should represent all the cases and controls represent the community from which they are drawn. Notice if the specific groups within the sample (level of instruction, employment, origin, etc) are proportionally represented. Assess if the cases and the controls come for the same population and if they come of a consecutive series or if there is any kind of selection involved.					
	Was the sample size estimated? Assess if the estimation has been made for employment, place of residence, level of instruction					
	The subjects who agreed to participate and the ones who actually participate affect the extrapolation of results to the population of interest? Analyze differences and similarities between the population of interest and the studied one taking into account the spatial and temporal context (e.g. prevalence of exposition), inclusion criteria, definition and measurement of the exposition and the outcome, confidence on the estimations, etc.					
	Summarizing, an adequate sample, similar to the population of interest minimizes the selection bias					

***Methods for measuring exposure and outcome variables: intervention /exposition, outcome, confounding o modifier variable** (In comparative studies the methods for measuring must be the same for both groups)	*The main variables have adequate conceptual (theoretical) and operational (measurement scale classification system, diagnostic criteria, etc.) definitions?* Clear definition of the cases and the controls. The state of illness of the patients in the cases group has been determined and validated in a reliable way.					
	The instruments used to measure the principal variables have validity and reliability which are known and adequate? Assess if studies which evaluated these instruments are correctly cited. For questionnaires originally designed for a determined culture or language, assess if they have been adapted to where they are going to be used. Evaluate if data was obtained from direct measurements, survey of medical records or self report events. Take into account if inter- and intra- observer variability have been addressed					
	The techniques used to measure the main variables are adequately described? Are appropriate? In case of comparative studies, are the same for both groups? Take into consideration the possibility of recall bias (the cases remember better the expositions than the controls) or interviewer bias. If the interviewer training was necessary, noted if it has been done.					
	Are there co-interventions or co-exposures likely to have biased the results?					
	The results were evaluated blindly or through objective criteria?					
	The follow up was long enough to allow observe the outcomes?					
	Was there quality control of the primary data?					
	Summarizing, the measurement and survey of the principal variables have been made in an appropriate way. The risk of information bias has been minimized					
***Methods for confounders control**	*Have been taken into account the potential main confounders in the design and analysis of the study?* In the design, variables associated with the studied problem should be taken into account; and in the analysis, estimation of the primary outcome should be stratified or adjusted by these variables (multivariate analysis). Notice if raw and adjusted measures have been reported, if the variables by which the results were adjusted are indicated, and if justification of which included (or not) in the analysis. Consider whether there were deviations in the protocol and the reasons of why.					

	Summarizing, all known confounding variables have been gathered and the analysis has been adjusted by them.					
***Comparability among groups.** (If there is no comparison between groups, the answer must be "not applicable" in each statement in this category)	*Are the source populations of the individuals in each group and the periods of recruitment similar?* (sociodemographic variables, prognostic factors or other variables who might alter the results). *According to the selection process, both populations have similar characteristics, are comparable throughout except for the study factors (cases and controls definition)?.* In cohort studies should assess whether patients were recruited at a similar point in the evolution of the disease.					
	Were the same strategies and measurement techniques used in both groups? Were the same variables measured?					
	The participant losses (missing data, dropouts, emigration, etc) affect the groups in different ways? The differences between groups concerning the losses must be less than 20%. Non-response rate and the reasons of non response are similar. A high non-response rate could be accepted if there is evidence that the groups remain similar. In case-control studies assess if there overmatching has been existed.					
	Summarizing, the groups are comparable; the possibility of selection bias has been minimized.					
MINOR CRITERIA						
Statistical Methods (excluding control of confounding)	*The statistical analysis was determined at the onset of the study?* (not post-hoc).					
	The statistical tests are appropriate? The allocation unit (community, institution, group, cluster, individuals) and the unit of analysis match? In case of mismatch an adjustment has been made? Are the methods appropriate to any anomalous data distributions (eg skewed, multimodal?) Was there adequate adjustment for any differences in length of follow-up (prospective studies) or time between the intervention and outcome (case-controls)?					
	In the analysis, losses of participants, missing data or other design defects (e.g. different probabilities of selection) were adequately treated? In certain cases the exclusion of cases to allow perform sensitivity analysis would be appropriate. An intention to treat analysis would be suitable for prospective comparative studies. Losses of more than 20% in short follow-up (less than a year) and more than 30% in larger periods should be considered high. In any case losses shouldn't affect the minimum sample size required to perform the statistical analysis.					
	Does reporting address parameter estimate uncertainty (ie with SEs or CIs)?					

	Was there a dose-response relationship between the intervention/exposition and the outcome shown?					
	Summarizing, the analysis is appropriate.					
Conflict of interest	***Summarizing, the conflict of interest does not condition the results or the conclusions of the study.***					

*Around half of the checklists included what we regard as the three most fundamental domains of appropriate selection of participants, appropriate measurement of variables and appropriate control of confounding.

#**Risk of Bias**
VH (Very High Risk of Bias): clearly indicates bias in all major domains
H (High Risk of Bias) clearly indicates bias in some of the major domains;
M (Moderate Risk of Bias) suggests potential bias in each domain;
L (Low Risk of Bias) clearly excludes bias in each domain;
? (Doubtful Risk of Bias) suggests doubts about potential bias in each domain