A Checklist for Discussants
Macartan Humphreys

Theory
- Is the theory internally consistent?
- Is it consistent with past literature and findings?
- Is it novel or surprising?
- Are elements that are excluded or simplified plausibly unimportant for the outcomes?
- Is the theory general or specific? Are there more general theories on which this theory could draw or contribute?

From Theory to Hypotheses
- Is the theory really needed to generate the hypotheses?
- Does the theory generate more hypotheses than considered?
- Are the hypotheses really implied by the theory? Or are there ambiguities arising from say non-monotonicities or multiple equilibria?
- Does the theory specify mechanisms?
- Does the theory suggest heterogeneous effects?

Hypotheses
- Are the hypotheses complex? (eg in fact 2 or 3 hypotheses bundled together)
- Are the hypotheses falsifiable?
- Are mechanisms implied by the hypotheses?

Evidence I: Design
- External validity: is the population examined representative of the larger population of interest?
- External validity: Are the conditions under which they are examined consistent with the conditions of interest?
- Measure validity: Do the measures capture the objects specified by the theory?
- Consistency: Is the empirical model used consistent with the theory?
- Mechanisms: Are mechanisms tested? How are they identified?
- Replicability: Has the study been done in a way that it can be replicated?
- Interpretation: Do the results admit rival interpretations?

Evidence II: Analysis and Testing
- Identification: are there concerns with reverse causality?
- Identification: are there concerns of omitted variable bias?
- Identification: does the model control for pre treatment variables only? Does it control or does it match?
- Identification: Are poorly identified claims flagged as such?
• Robustness: Are results robust to changes in the model, to subsetting the data, to changing the period of measurement or of analysis, to the addition or exclusion of plausible controls?
• Standard errors: does the calculation of test statistics make use of the design? Do standard errors take account of plausibly clustering structures/differences in levels?
• Presentation: Are the results presented in an intelligible way? Eg using fitted values or graphs? How can this be improved?
• Interpretation: Can no evidence of effect be interpreted as evidence of only weak effects?

Evidence III: Other sources of bias
• Fishing: were hypotheses generated prior to testing? Was any training data separated from test data?
• Measurement error: is error from sampling, case selection, or missing data plausibly correlated with outcomes?
• Spillovers / Contamination: Is it plausible that outcomes in control units were altered because of the treatment received by the treated?
• Compliance: Did the treated really get treatment? Did the controls really not?
• Hawthorne effects: Are subjects modifying behavior simply because they know they are under study?
• Measurement: Is treatment the only systematic difference between treatment and control or are there differences in how items were measured?
• Implications of Bias: Are any sources of bias likely to work for or against the hypothesis tested?

Policy Implications
• Do the policy implications really follow from the results?
• If implemented would the policy changes have effects other thank those specified by the research?
• Have the policy claims been tested directly?
• Is the author overselling or underselling the findings?