

## Appendix III: Sensitivity analysis

In order to investigate potential sources of observed heterogeneity in primary outcomes, we performed several subgroup and meta-regression analyses provided enough information was available.

For sex outcome severe disease, the first subgroup analysis included studies with quality scores 7 or above. This allows having only high-quality studies in the meta-analysis. Although the  $I^2$  statistics dropped to below 1% (from 15.2%), the effect size remained unaffected (RR 1.15, 95%CI 1.09 to 1.22), see Figure A1. As an additional analysis, we partitioned studies based on whether critical condition of severity was upon hospitalization or developed during follow-up. The former showed a slight increase (RR 1.27, 95%CI 1.12 to 1.44 – Figure A2) while the latter a slight decrease (RR 1.11, 95%CI 1.04 to 1.19 – Figure A3). However, both were fairly close to that of base analysis (RR 1.18, 95%CI 1.10 to 1.27). Finally, we performed meta-regression on study size, total quality score, study duration and study start date, but none were significant.

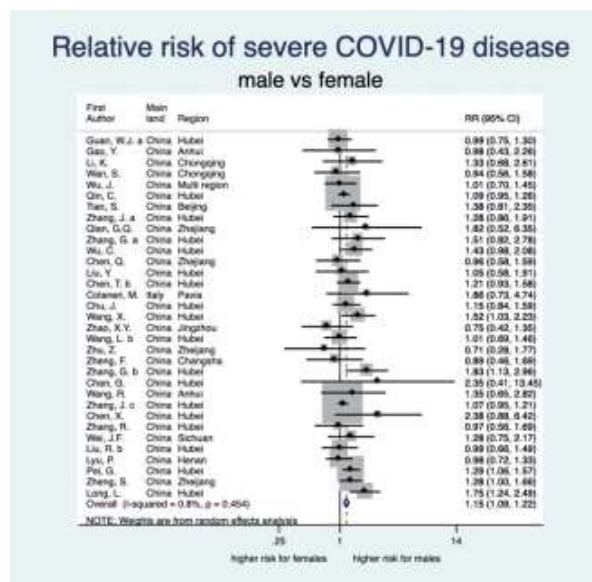


Figure A1

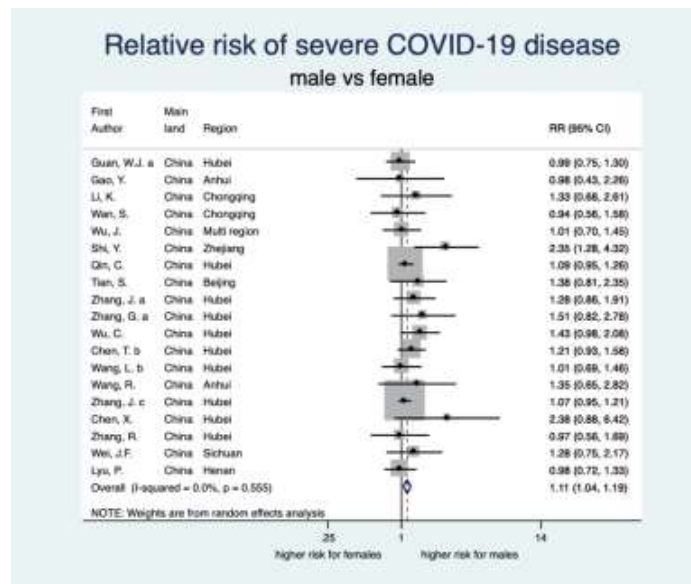


Figure A2

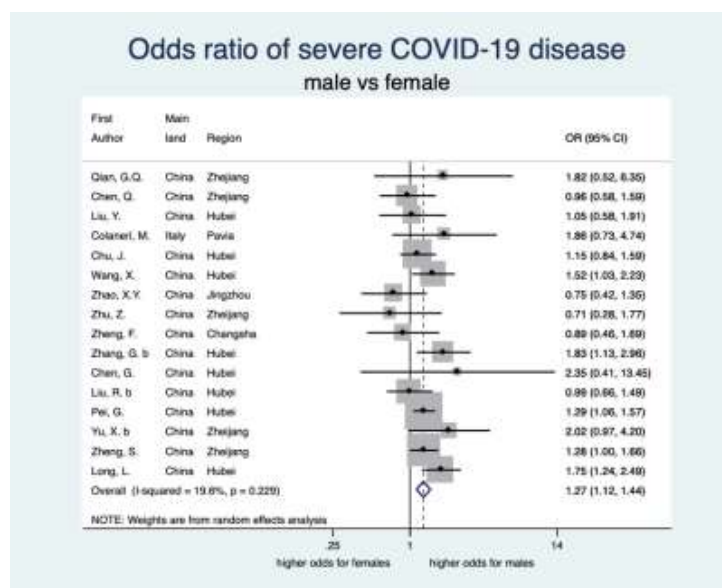


Figure A3

For sex outcome ICU admission, we conducted a subgroup analysis based on geographical location (Asia versus outside Asia), but the overall conclusion remained the same (RR 1.33, 95%CI 0.93 to 1.91 and RR 1.47, 95%CI 1.14 to 1.90 for Asia and outside Asia, respectively), see Figure A4. There was also no evidence for the effect of study size, total quality score, study duration and study start date from meta-regression.

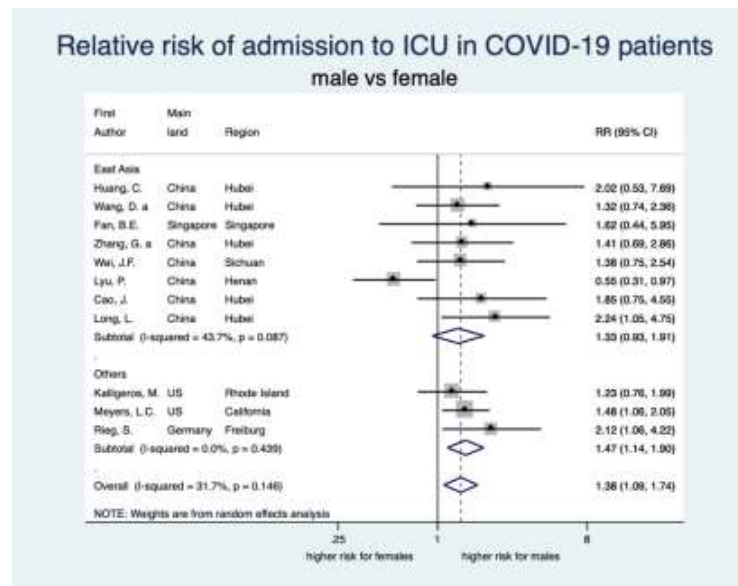


Figure A4

For sex outcome death, we also conducted a subgroup analysis based on geographical location (east Asia versus outside east Asia). In the group of east Asia, the effect size was substantially increased (RR 1.8, 95%CI: 1.32 to 2.46), while it largely dropped to RR 1.06, 95%CI: 0.93 to 1.22 in the group of outside east Asia, which consists of only 3 studies (see, Figure A5). The results from meta-regression on study start date revealed that this factor can explain about 40% of heterogeneity, see Table 1.

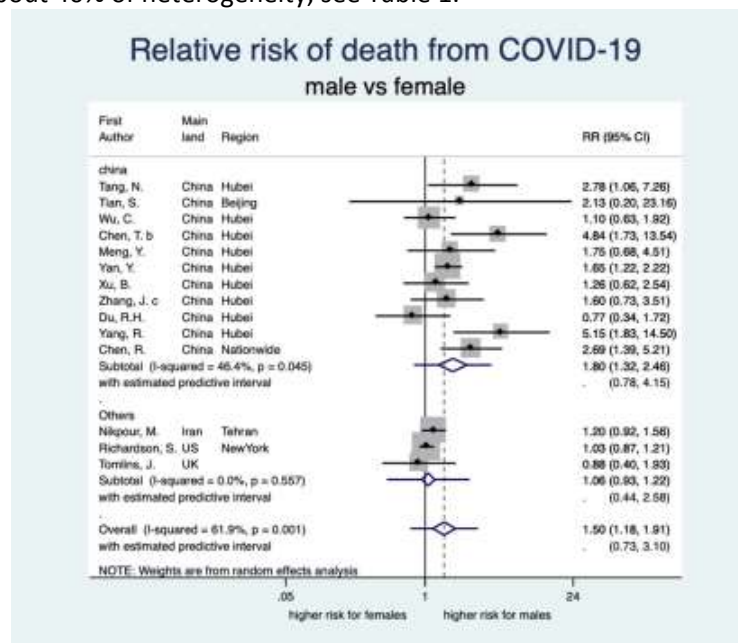


Figure A5

