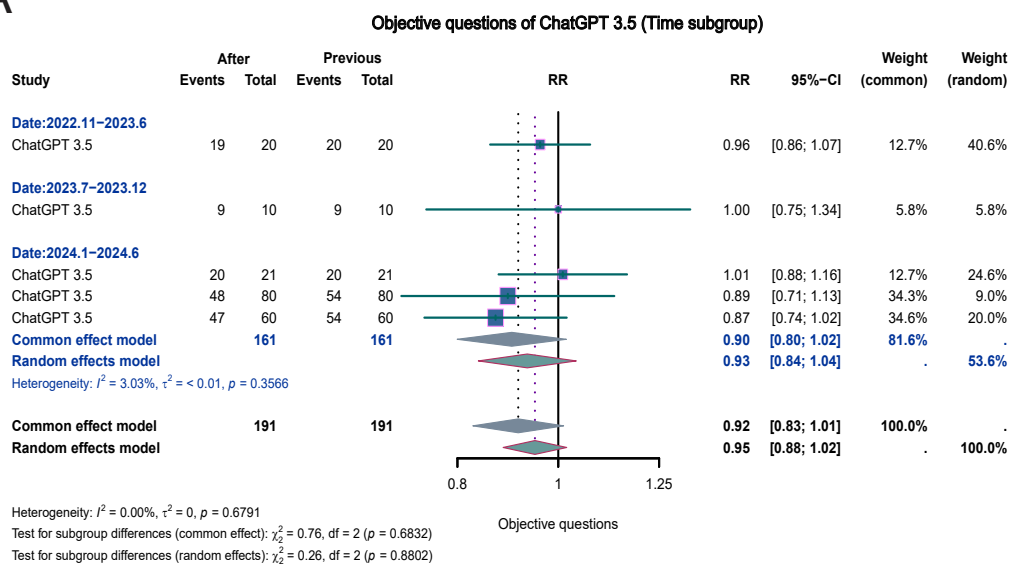
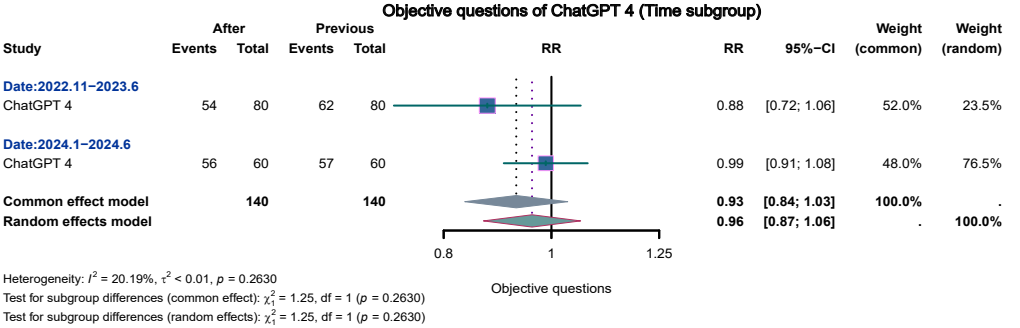
A


Objective questions of ChatGPT 3.5 (Time subgroup)
Study
After
Events
Total
Previous
Events
Total
RR
RR
95%−CI
Weight (common)
Weight (random)
Date:2022.11−2023.6
ChatGPT 3.5 19 20 20 20 0.96 [0.86; 1.07] 12.7% 40.6%
Date:2023.7−2023.12
ChatGPT 3.5 9 10 9 10 1.00 [0.75; 1.34] 5.8% 5.8%
Date:2024.1−2024.6
ChatGPT 3.5 20 21 20 21 1.01 [0.88; 1.16] 12.7% 24.6%
ChatGPT 3.5 48 80 54 80 0.89 [0.71; 1.13] 34.3% 9.0%
ChatGPT 3.5 47 60 54 60 0.87 [0.74; 1.02] 34.6% 20.0%
Common effect model 161 161 0.90 [0.80; 1.02] 81.6% .
Random effects model 0.93 [0.84; 1.04] . 53.6%
Heterogeneity: I^2 = 3.03%, τ^2 = < 0.01, p = 0.3566
Common effect model 191 191 0.92 [0.83; 1.01] 100.0% .
Random effects model 0.95 [0.88; 1.02] . 100.0%
0.8
1
1.25
Objective questions
Heterogeneity: I^2 = 0.00%, τ^2 = 0, p = 0.6791
Test for subgroup differences (common effect): χ^2_2 = 0.76, df = 2 (p = 0.6832)
Test for subgroup differences (random effects): χ^2_2 = 0.26, df = 2 (p = 0.8802)

B


Objective questions of ChatGPT 4 (Time subgroup)
Study
After
Events
Total
Previous
Events
Total
RR
RR
95%−CI
Weight (common)
Weight (random)
Date:2022.11−2023.6
ChatGPT 4 54 80 62 80 0.88 [0.72; 1.06] 52.0% 23.5%
Date:2024.1−2024.6
ChatGPT 4 56 60 57 60 0.99 [0.91; 1.08] 48.0% 76.5%
Common effect model 140 140 0.93 [0.84; 1.03] 100.0% .
Random effects model 0.96 [0.87; 1.06] . 100.0%
0.8
1
1.25
Objective questions
Heterogeneity: I^2 = 20.19%, τ^2 < 0.01, p = 0.2630
Test for subgroup differences (common effect): χ^2_1 = 1.25, df = 1 (p = 0.2630)
Test for subgroup differences (random effects): χ^2_1 = 1.25, df = 1 (p = 0.2630)