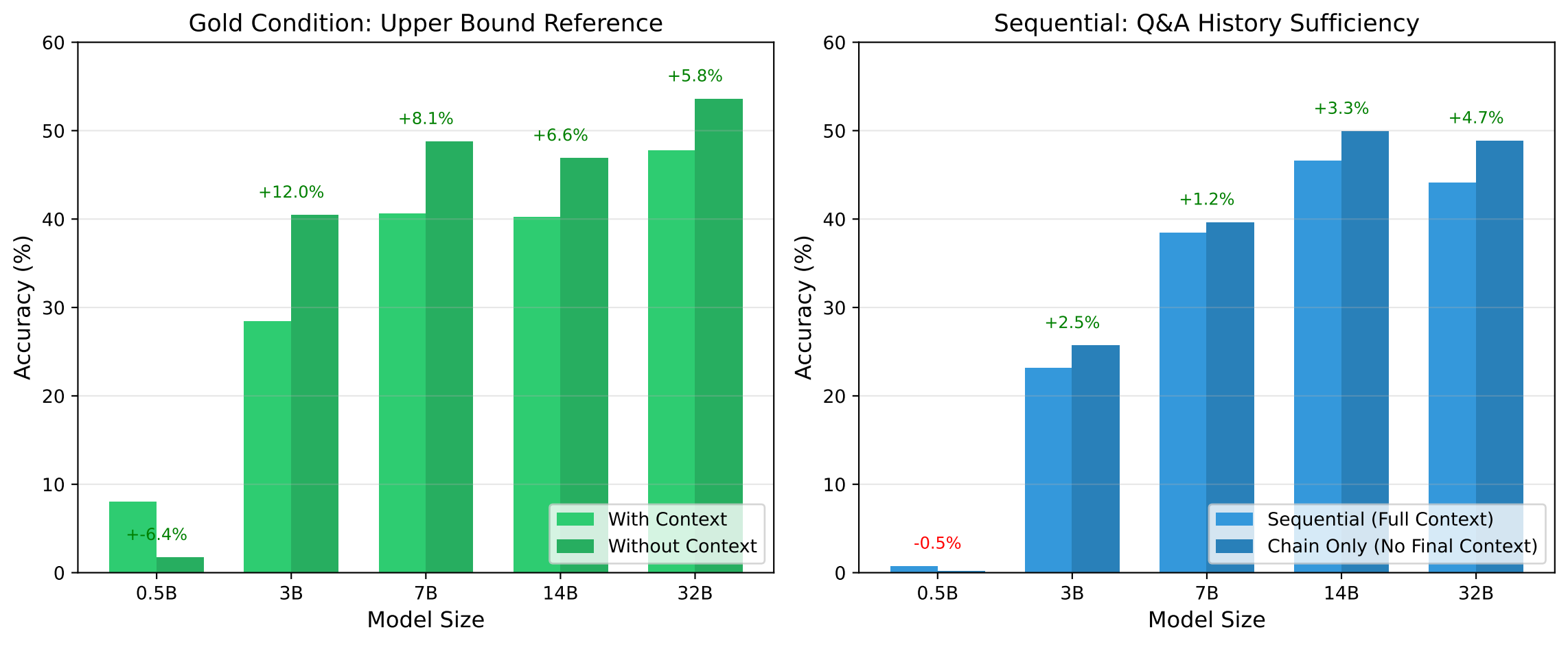
Gold Condition: Upper Bound Reference
Accuracy (%)
Model Size
0.5B
3B
7B
14B
32B
+-6.4%
+12.0%
+8.1%
+6.6%
+5.8%
With Context
Without Context
Sequential: Q&A History Sufficiency
Accuracy (%)
Model Size
-0.5%
+2.5%
+1.2%
+3.3%
+4.7%
Sequential (Full Context)
Chain Only (No Final Context)