

Supp. Fig.1  $\Delta$  Accuracy (v4 - baseline, pp) by question type  $\times$  subdomain