OBJECTIVES: The quality-adjusted time without symptoms and toxicity (Q-TWIST) methodology was used to assess the effectiveness of treatments (prolonged vs. progression-free) and costs (toxicity) of oncology therapies. This study was conducted to systematically review and quantitatively summarize published Q-TWIST assessments of cancer treatments. METHODS: A systematic search and review was conducted using PRISMA guidelines to identify original studies reporting the use of Q-TWIST—including time with toxicity (TOX), time before disease progression without toxicity (TWIST), and time in relapse after disease progression (REL)—for all oncology treatment groups, as available. Utilities for Q-TWIST were also captured; when a base case for utilities was not identified in a study, the following was assumed: u(TWIST) = 1, u(REL) = 0.5, and u(TOX) = 0.5. The relative gain in Q-TWIST for active treatment arms was calculated as the difference in Q-TWIST divided by mean overall survival of control arm. Relative gains ≥ 10% were considered to be a clinically important and clearly clinically important difference, respectively. RESULTS: Upon review of 84 initially identified articles, 39 were excluded for reasons including: absence of Q-TWIST with the LEAPS total score and productivity subscale change (F = 11, n = 22), not oncology-related (n = 13), other reasons (n = 4). Forty-five studies were included and reported a total of 69 Q-TWIST comparisons across 10 cancers. The most commonly used utilities for Q-TWIST calculation were u(TWIST) = 1, u(REL) = 0.5, and u(TOX) = 0.5 (n = 28, 32.8% of active treatment arms). The mean relative gain (0.05 ± 0.08). Other countries’ scoring algorithms are based on smaller sample sizes and there are considerable uncertainty (for comparison, the MID for EQ-5D utilities was 0.05 to 0.08). Other countries’ scoring algorithms are based on smaller sample sizes and there are considerable uncertainty (for comparison, the MID for EQ-5D utilities was 0.05 to 0.08). Other countries’ scoring algorithms are based on smaller sample sizes and there are considerable uncertainty (for comparison, the MID for EQ-5D utilities was 0.05 to 0.08). Other countries’ scoring algorithms are based on smaller sample sizes and there are considerable uncertainty (for comparison, the MID for EQ-5D utilities was 0.05 to 0.08). Other countries’ scoring algorithms are based on smaller sample sizes and there are considerable uncertainty (for comparison, the MID for EQ-5D utilities was 0.05 to 0.08). Other countries’ scoring algorithms are based on smaller sample sizes and there are considerable uncertainty (for comparison, the MID for EQ-5D utilities was 0.05 to 0.08). Other countries’ scoring algorithms are based on smaller sample sizes and there are considerable uncertainty (for comparison, the MID for EQ-5D utilities was 0.05 to 0.08). Other countries’ scoring algorithms are based on smaller sample sizes and there are considerable uncertainty (for comparison, the MID for EQ-5D utilities was 0.05 to 0.08). Other countries’ scoring algorithms are based on smaller sample sizes and there are considerable uncertainty (for comparison, the MID for EQ-5D utilities was 0.05 to 0.08). Other countries’ scoring algorithms are based on smaller sample sizes and there are considerable uncertainty (for comparison, the MID for EQ-5D utilities was 0.05 to 0.08). Other countries’ scoring algorithms are based on smaller sample sizes and there are considerable uncertainty (for comparison, the MID for EQ-5D utilities was 0.05 to 0.08). Other countries’ scoring algorithms are based on smaller sample sizes and there are considerable uncertainty (for comparison, the MID for EQ-5D utilities was 0.05 to 0.08). Other countries’ scoring algorithms are based on smaller sample sizes and there are considerable uncertainty (for comparison, the MID for EQ-5D utilities was 0.05 to 0.08). Other countries’ scoring algorithms are based on smaller sample sizes and there are considerable uncertainty (for comparison, the MID for EQ-5D utilities was 0.05 to 0.08). Other countries’ scoring algorithms are based on smaller sample sizes and there are considerable uncertainty (for comparison, the MID for EQ-5D utilities was 0.05 to 0.08).