Mapping to EQ-5D Utility in Knee Osteoarthritis: A Comparative Evaluation of WOMAC and SF-36 Algorithms in the STEP 9 Trial

PCR6

Vagia Daki[1]; Karthik Ramakrishnan[2]; Konstantinos Soulanis[1]; Ines Guerra[3]; Theodora Oikonomidi[1]; Inger Smith[4]

Aim

- This research aimed to derive estimates of the EuroQol Three Dimension (EQ-5D-3L) utility index scores for the STEP 9 trial, by means of mapping patient-reported outcomes (PROs), assessed in the STEP 9 trial, to EQ-5D-3L utilities.
- The objective was to evaluate the face validity of EQ-5D-3L utility estimates derived via mapping from Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) and 36-Item Short Form Health Survey (SF-36) measured in the STEP 9 trial.

Introduction

- The STEP 9 trial(1) evaluated subcutaneous semaglutide (2.4 mg) plus dietary interventions and physical exercise (D&E) versus D&E alone in patients with knee osteoarthritis (OA) and obesity.
- The trial collected PROs via WOMAC and SF-36 questionnaires but lacked direct measurement of EQ-5D-3L utility data, requiring utility mapping to derive EQ-5D-3L values needed for conducting cost-effectiveness analysis (CEA) of semaglutide in knee OA.
- Rigorous evaluation and transparent selection of mapping methods was undertaken to strengthen credibility of utility estimation for the CEA.

Methods

- A stepwise approach was followed to identify published WOMAC or SF-36 mapping algorithms suitable for estimating ED-5D-3L utilities.
  - A search of the Health Economics Research Centre (HERC) database (2) was conducted to identify mapping algorithms.
    - WOMAC and SF-36 based algorithms were evaluated according to NICE guidelines(3).
    - Evaluation criteria included population comparability to the STEP 9 trial, predictive accuracy of mapping algorithm, and results of external validation (i.e., performance assessed in a dataset not used for training).
  - A targeted desk search of NICE appraisals explored use of WOMAC and SF-36 mapping in HTA submissions of knee disorders and obesity.
  - The most appropriate WOMAC and SF-36 mapping algorithm was selected to derive EQ-5D-3L utilities in the STEP-9 trial.
  - A targeted literature review (TLR) identified studies reporting changes in WOMAC and SF-36 scores from baseline and corresponding impact on EQ-5D-3L utilities.
  - Changes in mapped utility estimates and baseline values in the STEP 9 trial were validated against published data identified via the TLR.

Results

Overview of STEP 9 trial (n=407)

Table 1: Baseline characteristics of the STEP 9 trial population

Study arm	Age	BMI	WOMAC pain score	SF-36 PCS score	SF-36 MCS score
SEMA+D&E	56	40.5	72.8	32.8	51.1
D&E	56	40.0	67.2	33.8	52.4

BMI, Body-mass index; D&E, Dietary interventions and physical exercise; MCS, Mental component summary; PCS, Physical component summary; Values presented are means.

WOMAC to EQ-5D Mapping Algorithms Review

- Six mapping algorithms from WOMAC to EQ-5D were identified (Table 2).
  - Three studies (Ayala 2021(4), Bilbao 2020(5), Price 2019(6)) did not externally validate the suggested algorithm.
  - Most algorithms performed poorly in predicting utilities in severe health states (Ayala 2021, Bilbao 2020, Xie 2010(7), Barton 2008(8)).
  - A weak association between the WOMAC stiffness subscale and EQ-5D was found in three studies (Ayala 2021, Bilbao 2020, Wailoo 2014(9)).
  - Several differences were noted between these studies and the STEP 9 trial based on baseline age, BMI (Data Supplement, Table 1), and disease severity (as per WOMAC pain scores; Table 2).

Table 2: Overview of the WOMAC-to-EQ-5D mapping studies

Study	Population	Age	Predictive accuracy	External validation	WOMAC pain score
Wailoo 2014	Knee/hip OA	69.1	High	Yes	56.9
Xie 2010	Knee OA	66.5	Moderate to High	Yes	33.2
Barton 2008	Knee pain	NR	Moderate	Yes	38.8
Ayala 2021	Knee/hip OA	69.6	Moderate	No	46.4
Bilbao 2020	Knee/hip OA	69.8	Moderate	No	46.4
Price 2019	Knee OA	69.1	Moderate	No	42.5

NR, Not reported; Accuracy ranking was based on Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). Green, yellow, and pink indicate high, moderate, and low appropriateness for STEP 9, respectively; Values presented are means.

WOMAC to EQ-5D Mapping Algorithms Review

- Wailoo 2014 was deemed most appropriate to map STEP 9 data based on evaluation criteria and closest alignment to the STEP 9 trial.
- No NICE submissions employing WOMAC mapping were identified through targeted desk research.

SF-36 to EQ-5D Mapping Algorithms Review

- Studies mapping SF-36 scores to EQ-5D-3L utilities in Knee OA are lacking.
- Four SF-36 to EQ-5D mapping algorithms were identified in other populations: Rowen 2009(10), Ara-Brazier 2008(11), Maund 2012(12), and Kim 2014(13) (Table 3).
- Only Maund 2012 employed a flexible statistical framework appropriate for the skewed utility distribution; the others used linear regression.
- Rowen 2009 and Maund 2012 lacked external validation.
- Ara-Brazier 2008 and Kim 2014 reported weak correlations between EQ-5D and SF-36 subscales (role physical and vitality).
- Most algorithms showed limited accuracy in predicting EQ-5D utilities in severe health states.

Table 3: Overview of population and key methodological features in the SF-36-to-EQ-5D mapping studies

Study	Population	Age	Predictive accuracy	External validation	SF-36 PCS score	SF-36 MCS score
Rowen 2009	Various conditions	58.1	Moderate	No	38.3	44.9
Ara-Brazier 2008	Various conditions	52.0	High	Yes	NR	NR
Maund 2012	Rotator cuff disease	Range: 55-59	Low	No	36.0 / 39.0	44.7/ 46.5
Kim 2014	General population	56.9	High	Yes	44.7	43.9

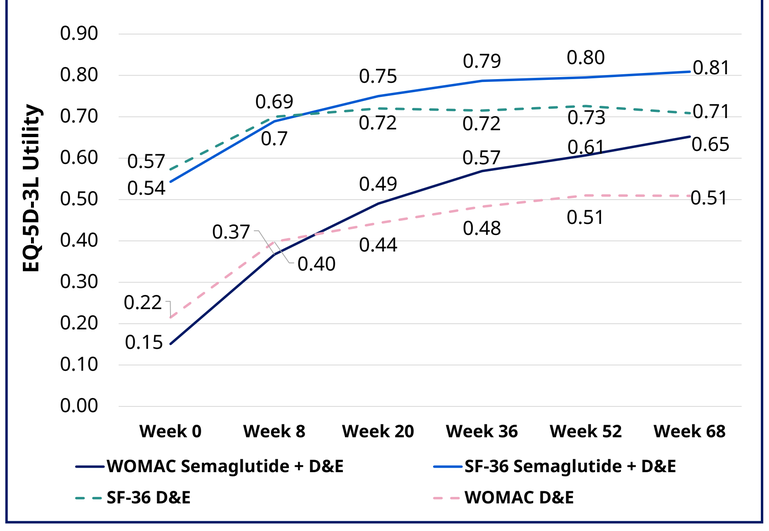
Accuracy ranking was based on Mean Absolute Error (MAE) and coefficient of determination (R^2). Green, yellow, and pink indicate high, moderate, and low appropriateness for STEP 9, respectively; Values presented are means unless otherwise indicated.

- Rowen 2009 was selected for implementation in the STEP 9 trial due to:
  - Physical Component Summary (PCS) and Mental Component Summary (MCS) scores more closely aligned with STEP 9 than Ara-Brazier 2008 and Kim 2014 (Table 3).
  - Similar consistency observed across individual SF-36 domains (Data Supplement, Figure 1).
  - Stronger predictive performance (R^2 = 0.70 vs. 0.40 in Maund 2012).
- The desk research identified one NICE appraisal in knee OA that used the Rowen 2009 mapping algorithm, which was accepted.

EQ-5D-3L Utilities Mapped from WOMAC and SF-36 in STEP 9 trial

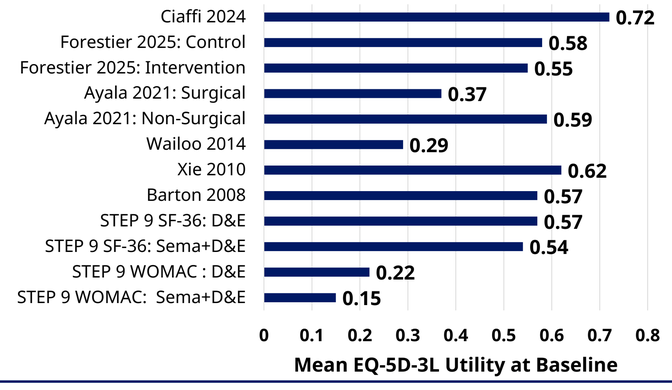
- QoL data from the STEP 9 trial, collected via WOMAC and SF-36, were mapped to EQ-5D-3L index scores using Wailoo 2014 and Rowen 2009 algorithms, respectively (Figure 1).
- Baseline utility values were much lower with WOMAC mapping compared to SF-36 mapping (Figure 1):
  - WOMAC mapping: 0.15 for semaglutide + D&E and 0.22 for D&E alone.
  - SF-36 mapping: 0.54 for semaglutide + D&E and 0.57 for D&E alone.
- Both semaglutide and D&E arms showed utility gains over time, with a linear increase observed from Week 8 to 68 (Figure 1).
  - SF-36 mapping yielded smaller utility improvements than WOMAC from baseline to Week 68;
    - For semaglutide + D&E, the increase was 0.27 with SF-36 mapping vs. 0.50 with WOMAC mapping; for D&E alone, it was 0.14 vs. 0.29, respectively.

Figure 1: EQ-5D utilities mapped from WOMAC and SF-36 in the STEP 9 trial, over time (Week 0-68)



- Baseline EQ-5D utilities from the STEP 9 trial (Figure 2) were questionably low when derived via WOMAC mapping. In contrast, SF-36-based utilities aligned with baseline ranges (0.29–0.72) reported in studies presenting WOMAC mapping algorithms and those identified through the TLR.

Figure 2: Comparison of baseline EQ-5D utilities from the STEP 9 trial, WOMAC mapping, and TLR studies



TLR: Impact of WOMAC Pain score on EQ-5D Utility in Knee OA

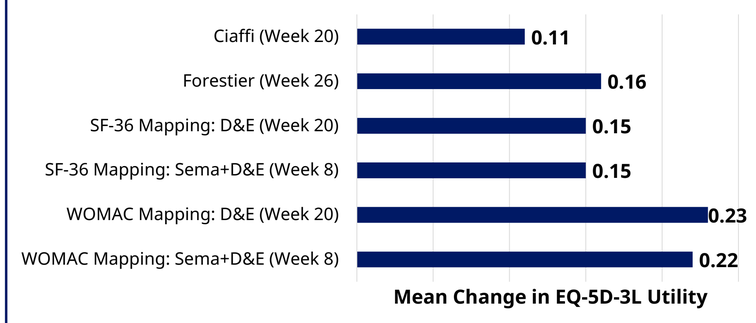
- The TLR identified three studies. Two studies including one RCT(14) and one pilot interventional study(15) were analysed; the remaining study(16) did not report baseline QoL scores, limiting comparison to STEP 9 trial.
- Compared to the STEP 9 trial, the included studies reported lower mean BMI and better WOMAC pain and QoL scores at baseline (Table 4).
- Despite population differences to STEP 9, a 20-point reduction in WOMAC pain score yielded utility gains of 0.11-0.16 (Figure 3).
- In the STEP 9 trial, a similar WOMAC pain reduction (~20 points) was observed at Week 8 with semaglutide + D&E and at Week 20 with D&E alone, corresponding to utility gains of 0.22 and 0.23 (WOMAC mapping) and 0.15 (SF-36 mapping) for both arms (Figure 3).
- Utility gains in the STEP 9 trial estimated using SF-36 mapping were more aligned with the TLR studies.

Table 4: Baseline characteristics of studies identified in the TLR on EQ-5D-3L utilities in knee OA

Study	Arm	BMI	EQ-5D-3L utility	WOMAC pain score	SF-36 PCS score	SF-36 MCS score
Forestier 2025 (n=74; 71)	3-week spa therapy	29.8	0.55	57.7	NR	NR
	Control group	30.1	0.58	52.0	NR	NR
Ciaffi 2024 (n=16)	20-week VLCKD	40.0	0.72	40.0	46.2	53.6

NR, Not reported; VLCKD, Very low-calorie ketogenic diet; Values presented are means.

Figure 3: Utility impact of 20-point WOMAC pain reduction in the STEP 9 trial and the TLR studies



STEP 9 data reflect Weeks 0–8 for semaglutide + D&E and Weeks 0–20 for D&E alone, corresponding to a 20-point reduction in WOMAC pain.

Discussion

- Baseline EQ-5D utilities and improvements over time in the STEP 9 trial, derived using the SF-36(10) mapping versus WOMAC mapping(9) method appear more plausible and aligned with identified studies in this review.
- Mapping EQ-5D utilities from WOMAC scores may be limited due to weak associations—particularly with the stiffness subscale—and the absence of mental health domains.
- SF-36 to EQ-5D mapping has precedent in NICE appraisals via the Rowen 2009 algorithm, whereas WOMAC mapping has not been used.
- Both mapping approaches show limitations for the STEP 9 trial, due to population heterogeneity and limited accuracy in predicting severe health states.

Conclusion

- Use of SF-36-based mapping in the STEP 9 trial, yielded EQ-5D-3L utility estimates that are more consistent with published literature than WOMAC-based mapping.
- Both mapping strategies have limitations due to:
  - Population heterogeneity compared to STEP 9 trial.
  - Reduced accuracy in predicting utilities in extreme or severe health states.
- Future research should focus on:
  - Improving utility mapping methodologies in Knee OA.
  - Validating algorithms in diverse Knee OA patient populations to enhance the reliability of utility estimates for CEA estimation and HTA decision-making.

[1]IQVIA, Athens, Greece; [2]Novo Nordisk A/S, Bengaluru, India, [3]IQVIA, London, United Kingdom; [4]White Box Health Economics Ltd, Brighton & Hove, UK

This trial was sponsored by Novo Nordisk and is registered with ClinicalTrials.gov (NCT05064735).
Presented at the ISPOR EU conference, 2025, Glasgow.

References:
(**1**) Bliddal et al. N Engl J Med. 2024:391;1573–83; (**2**) Dakin et al. HERC Database of Mapping Studies. 2023:9.0;1–45; (**3**) Wailoo et al. TSD 22. ScHARR, University of Sheffield. 2023:22;1–45; (**4**) Ayala et al. Value Health. 2021:24;874–83; (**5**) Bilbao et al. Value Health. 2020:23;379–87; (**6**) Price et al. Health Technol Assess. 2019:23;1; (**7**) Xie et al. Value Health. 2010:13;873–78; (**8**) Barton et al. Health Qual Life Outcomes. 2008:6;1–11; (**9**) Wailoo et al. Health Qual Life Outcomes. 2014:12;1–6; (**10**) Rowen et al. Health Qual Life Outcomes. 2009:7;1–9; (**11**) Ara et al. Value Health. 2008:11;1131–43; (**12**) Maund et al. NIHR Exec Summ. 2012:12;1–45; (**13**) Kim et al. Health Qual Life Outcomes. 2014:12;1–0; (**14**) Forestier et al. Int J Biometeorol. 2025:NA;1–10 ;(**15**) Ciaffi et al. Nutrients. 2024:16;3236; (**16**) Ateş et al. Acta Orthop Traumatol Turc. 2021:55(5);417–21;