Supplementary Material

Comparative Effectiveness of Biologic Classes in Clinical Practice: Month 12 Outcomes from the International Observational Psoriasis Study of Health Outcomes (PSoHO)

Saakshi Khattri[1], Álvaro González-Cantero[2,3], Burhan Engin[4], Sunil Dogra[5], Caroline A. Murphy[6], Christopher Schuster[6,7], Naoto Tsujimoto[8], Georgia Martimianaki[6], Anastasia Lampropoulou[6], Aya Alsharafi[6], Bruce Konicek[6], Felix Lauffer[9]

[1]Department of Dermatology, Icahn School of Medicine at Mount Sinai, New York City, USA.

[2]Department of Dermatology, Hospital Universitario Ramón y Cajal, IRYCIS, Colmenar Viejo km 9.100, 28034 Madrid, Spain.

[3]Faculty of Medicine, Universidad Francisco de Vitoria, 28223 Madrid, Spain.

[4]Department of Dermatology, Istanbul University-Cerrahpasa, Faculty of Medicine, Istanbul, Turkey.

[5]Department of Dermatology, Venereology, and Leprology, Postgraduate Institute of Medical Education and Research, Chandigarh, India.

[6]Eli Lilly and Company, Indianapolis, IN, USA.

[7]Department of Dermatology, Medical University of Vienna, Vienna, Austria.

[8] Eli Lilly and Company, Kobe, Japan.

[9]Department of Dermatology and Allergy, Technical University of Munich, 80802 Munich, Germany.

Corresponding Author: Caroline Murphy, Global Scientific Communications, Eli Lilly and Company, Island Hall, Eastgate Road, Eastgate Business Park, Little Island, Co. Cork, Ireland.

Email: caroline.murphy@lilly.com

Tables and Figures

Table 1: Baseline demographics of the US on-label patient population

	Overall (N=1674)	Anti-IL-17A/RA biologics (N=721)	Anti-IL-23 biologics (N=567)	Anti-IL-12/23 biologics (N=99)	Anti-TNFα (N=287)
Age, years	45.3 (13.6)	46.8 (13.7)	44.2 (13.2)*	46.1 (15.0)	43.6 (13.0)*
Male, n (%)	958 (57.2)	412 (57.1)	338 (59.6)	58 (58.6)	150 (52.3)
Weight (kg)	85.2 (21.0)	85.6 (20.5)	84.6 (21.9)	83.5 (18.3)	86.0 (21.4)
BMI, kg/m^2	29.1 (6.7)	29.2 (6.5)	29.0 (6.8)	28.2 (5.8)	29.4 (7.0)
Race- White, n(%)	1252 (74.8)	551 (76.4)	377 (66.5)*	80 (80.8)*	244 (85.0)**
Race- Asian, n(%)	225 (13.4)	91 (12.6)	123 (21.7)*	5 (5.1)**	6 (2.1)*
Race- Not Reported , n (%)	193 (11.5)	76 (10.5)	66 (11.6)	14 (14.1)	37 (12.9)
Median Disease Duration, Years (Q1, Q3)	14.4 (7.23, 24.02)	15.2 (7.1, 24.8)	14.6 (8.2, 23.9)	11.5 (5.8, 23.5)	13.0 (5.9, 22.8)
PASI Total Score	14.7 (8.7)	14.8 (8.7)	15.2 (9.5)	14.7 (8.0)	13.7 (7.2)
% BSA	21.8 (18.01)	21.9 (17.9)	21.4 (18.3)	23.4 (18.4)	21.8 (17.5)
sPGA, n (%), **Moderate** **Severe** **Very Severe**	829 (50.3%) 516 (31.3%) 71 (4.3%)	363 (51.1) 217 (30.6) 35 (4.9)	248 (44.5) 193 (34.6) 30 (5.4)	54 (55.1) 27 (27.6) 2 (2.0)	164 (58.2) 79 (28.0) 4 (1.4)
DLQI[a]	12.7 (7.8)	13.0 (7.8)	11.9 (7.7)**	12.5 (7.8)	13.5 (7.6)
Any Comorbidities Reported, n (%)	977 (58.4%)	435 (60.3)	323 (57.2)	61 (61.6)	158 (55.1)
Number of Current Comorbidities Reported [b]	1.5 (1.76)	1.6 (1.8)	1.4 (1.7)	1.7 (2.1)	1.3 (1.6)
Psoriatic Arthritis, n (%) [c]	388 (23.2)	207 (28.7)	103 (18.2)*	16 (16.2)**	62 (21.6)**
Nail Psoriasis, n (%) [d]	643 (38.5)	289 (40.1)	217 (38.4)	33 (33.7)	104 (36.2)
Any Previous Conventional Therapy, n (%)	1330 (79.5)	546 (75.8)	436 (76.9)	85 (85.9)**	263 (91.6)*
Any Previous Biologic Therapy, n (%) [e]	604 (36.1)	270 (37.5)	278 (49.0)*	28 (28.3)	28 (9.8)*

All results are expressed as mean (standard deviation) of all available data for that measure, unless otherwise indicated.
BMI, body mass index, BSA: body surface area, DLQI: Dermatology Life Quality Index, IL: interleukin, PASI: Psoriasis Area and Severity Index, PsA: psoriatic arthritis, Q1: quartile 1, Q3: quartile 3, sPGA: Static Physician Global Assessment, TNF: tumor necrosis factor
**P-value < 0.05 vs. Anti-IL-17A/RA (shaded in gray).
*P-value < 0.001 vs. Anti-IL-17A/RA (shaded in green).
[a]DLQI was measured on a 0–30 scale.
[b]Comorbidities were captured based on a predefined list.
[c]PsA diagnosis was recorded by the dermatologists based on the medical history and/or information provided by the patient.
[d]Recorded as a simple yes/no question (investigator assessed).
[e] Information about prior biologic use missing in 1 patient.

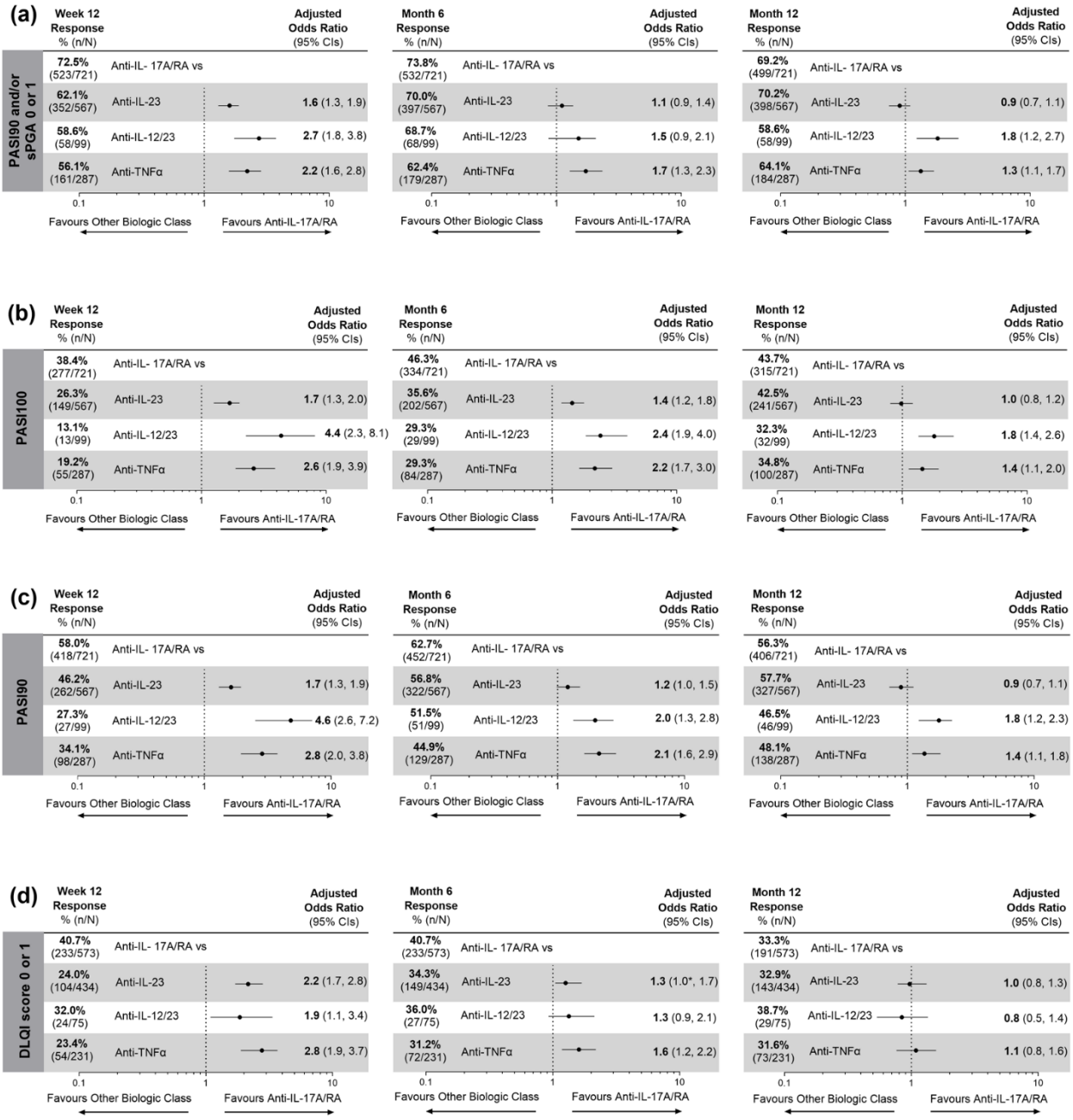

*Figure 1: Unadjusted response rates and comparative adjusted odds ratios for primary and secondary outcomes for anti-IL-17A/RA versus other drug classes for the United States on-label patient population. A) Primary outcome of PASI90 and/or sPGA score of 0 or 1 at week 12, month 6, and month 12. B) PASI100 at week 12, month 6, and month 12. C) PASI90 at week 12, month 6 ,and month 12. D) DLQI score of 0 or 1 at week 12, month 6, and month 12. DLQI outcome only included patients with an available DLQI score of 2 or greater at baseline. Results are statistically significant if 95% CIs of the odds ratios do not cross 1. For instances that lower CI shows 1.0, * denotes that the result is significant as lower CI is greater than 1.*

CI: confidence interval, DLQI: Dermatology Life Quality Index, IL: interleukin, n: number of patients who achieved the outcome, N: total number of patients in treatment group, PASI: Psoriasis Area and Severity Index, PASI90: 90% improvement in the PASI score, PASI100: 100% improvement in the PASI score, sPGA: Static Physician Global Assessment, TNF: tumor necrosis factor.

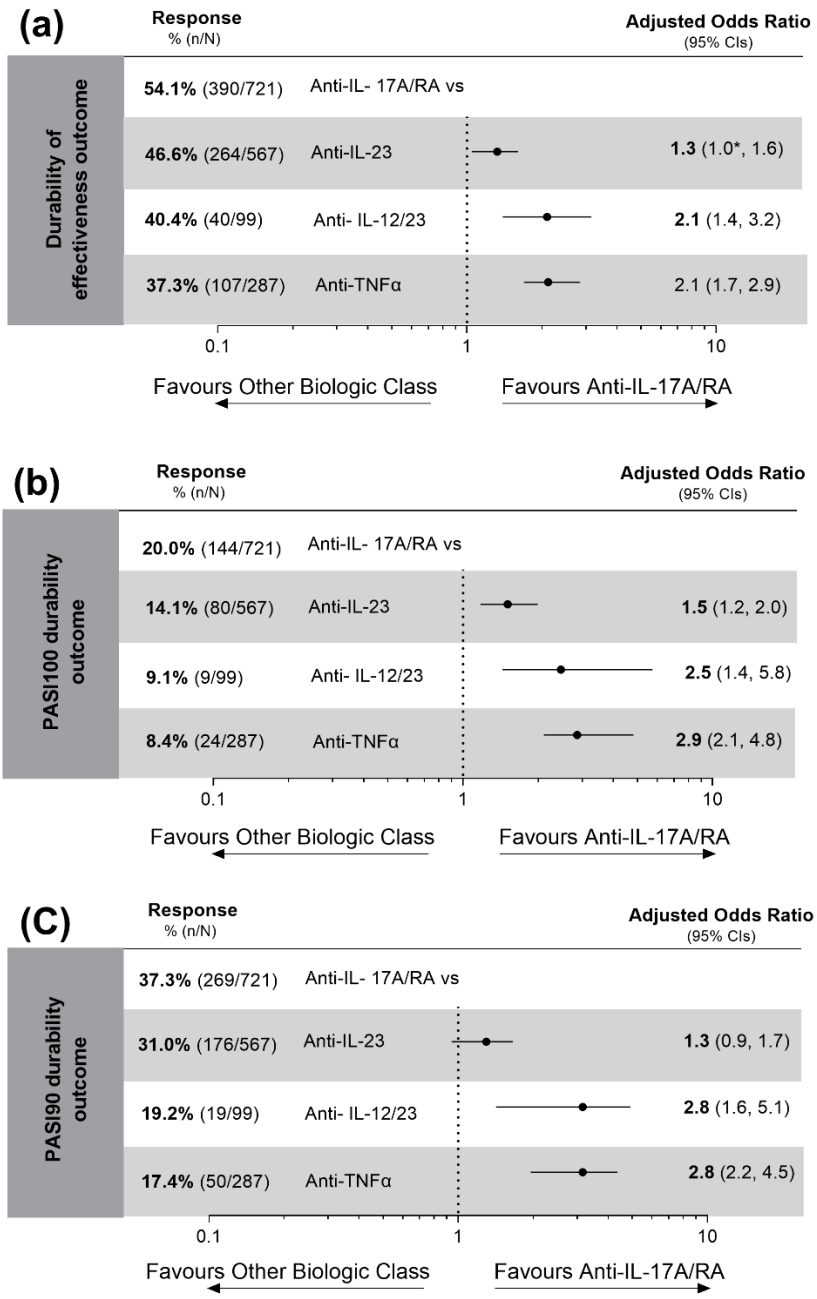

*Figure 2: Unadjusted response rates and comparative adjusted odds ratios for durability of effectiveness outcomes for anti-IL-17A/RA versus other drug classes for the United States on-label patient population. Outcomes include a A) durability of effectiveness outcome, whereby patients achieve at least PASI90 and/or sPGA score of 0 or 1 at week 12 and subsequently achieve either or both of the following outcomes at months 6 and 12: at least PASI75 or achieving an improvement in sPGA of 2 or more points from baseline. B) PASI100 durability outcome, whereby patients achieve a PASI100 score at week 12 and subsequently at both months 6 and 12. C) PASI90 durability outcome, whereby patients achieve a PASI90 score at week 12 and subsequently at both months 6 and 12. Results are statistically significant if 95% CIs of the odds ratios do not cross 1. For instances that lower CI shows 1.0, * denotes that the result is significant as lower CI is greater than 1.*

CI: confidence interval, IL: interleukin, n: number of patients who achieved the outcome, N: total number of patients in treatment group, PASI: Psoriasis Area and Severity Index, PASI75: 75% improvement in the PASI score, PASI90: 90% improvement in the PASI score, sPGA: Static Physician Global Assessment, TNF: tumor necrosis factor.

Statistical Appendix

Application of frequentist model averaging to real-world data

A comparative analysis utilizing PSoHO's real-world data was conducted in a machine learning-based frequentist model averaging (FMA) framework for causal inference, performing the same statistical methodology as previously described (1, 2). For the current analysis, an estimate of the average treatment effect (ATE) was obtained by the treatment and outcome models listed in Supplementary Appendix Table 2. The treatment models used either gradient boosting or logistic regression to generate a propensity score, which allowed for the balance of the treatment groups (3). Generalized linear models (GLM) (unadjusted or adjusted) and penalized regression (PR) models (lasso or elastic net) were among the outcome models (4, 5).

Table 2: Individual statistical methods used within the FMA analysis

Adjustment method	Specifications	Treatment Model	Outcome model
Weighting	IPTW	LR (covariates* selected based on PR) GBTM	1. Treatment only 2. Elastic Net with Treatment + all covariates* 3. Lasso with Treatment + all covariates*
Regression*	GLM, lasso or elastic net	NA	NA

GBTM: gradient boosted tree model, GLM: generalized linear models, IPTW: Inverse Probability Treatment weighting, FMA: frequentist model averaging, LR: logistic regression, NA: not applicable, PR: Penalized regression. *Variables used are the same as for the balancing score.

All analyses were conducted using SAS 9.4 and R Version 4.0.3 (R Foundation for Statistical Computing, Vienna, Austria) or later. The baseline covariates used to adjust all models are listed in Supplementary Appendix Table 3 (1).

Table 3: Covariates

At Baseline		
Patient demographics	**Selected comorbidities associated with psoriasis**	**Patient disease characteristics**
Age	Diabetes Mellitus	Number of Current Comorbidities Reported
Sex	Inflammatory bowel disease	Baseline PASI score
BMI (kg/m²)	Ankylosing spondylitis (AS) and Axial AS	Baseline sPGA score
Time Since Onset of Plaque Psoriasis (years)	Obesity	Baseline BSA
Biologic Naïve vs. Biologic Non-naïve-C	Overweight	Baseline Itch VAS score
Blood Relatives with Psoriasis	Dyslipidemia	Baseline Skin VAS score
Previous phototherapy	Diagnosis of Psoriatic Arthritis	Psoriasis visible on areas of:
Any Previous Non-Biologic Systemic Therapy	Rheumatic disease unrelated to PsO/PsA	• the face and/or neck
Currently Taken Any Concomitant Medications	Other comorbidities associated with psoriasis	• the scalp
Current Non-biologic Systemic Therapy		• the palms and/or soles
Current Conventional Therapy		• the genital area
Current Phototherapy		Baseline mNAPSI score
Current Topical Agent		Baseline WPAI-PSO Score
How often do you have six or more drinks on one occasion?		Baseline DLQI score
Smoking Status		Baseline HADS score (total/anxiety/depression)
		MARS Score
		Psoriasis Symptoms and Signs Diary (PSSD)

AS: Ankylosing spondylitis, BSA: Body Surface Area, DLQI: Dermatology Life Quality Index, HADS: Hospital Anxiety and Depression Scale, mNAPSI: modified Nail Psoriasis Severity Index, PASI: Psoriasis Area Severity Index; PsA: psoriatic arthritis, PsO: psoriasis, PSSD: Psoriasis Symptoms and Signs Diary, sPGA: static Physician's Global Assessment, VAS: visual analog scale, WPAI-PSO: Work Productivity and Activity Impairment Questionnaire - Psoriasis.

Constructing FMA estimates of treatment effect and associated confidence intervals

The uncertainty in FMA estimates was quantified using percentile bootstrap method. Specifically, B=100 bootstrap samples were extracted from the original data (For baseline characteristics, missing values were imputed as mean for continuous variables, as the most frequent category for categorical variables and for all the outcome variables non-responder imputation was used), with the entire longitudinal patient profile as the sampling unit. The data were sampled with replacement using 5-fold cross-validated FMA weights stratified by treatment arms. This ensures the same distribution of treatments across bootstrap samples and prevents allocating multiple replicates of the same subject (within a bootstrap sample) to different folds that would have compromised computing cross-validation prediction error (6).

For each bootstrapped sample $b = 1, .., B$ $(B = 100)$, all individual strategies applied on each set resulted in FMA estimates of treatment effect, logit $(\widehat{or}_b^{fma})$.

The final FMA estimate $(\widehat{or}^{fma})$ on bootstrap sample was determined as

$$\widehat{or}^{fma} = \exp\left(B^{-1}\sum_{b=1}^{B}\text{logit}(\widehat{or}_b^{fma})\right)$$

and the 95% confidence interval computed by exponentiating the 2.5% and 97.5% percentiles of the bootstrap distribution of $logit(\widehat{or}_b^{fma})$ $b = 1,..,B$. The confidence intervals for treatment effects by individual strategies were computed similarly.

Main analysis using non-responder imputation of missing data

As has previously detailed (1, 2), we applied non-responder imputation to missing outcomes for the main analysis (19.9% in the primary outcome at month 12). Furthermore, to assess the potential impact of unmeasured confounding and support the reported treatment estimates' robustness we calculated the E-value (1, 2, 7). The E-value for the PASI 90 durability outcome (NRI) for the comparison of the Anti-IL17A/RA cohort vs Anti-IL-23 cohort was 1.54, and the E-value for the lower confidence limit of the point estimate was 1.28. The E-value, shown in Supplementary Appendix Figure 3, indicates that an unmeasured confounder would need to have an association of RR > 1.54 with both the treatment selection and the outcome to shift the observed treatment estimate toward the null.

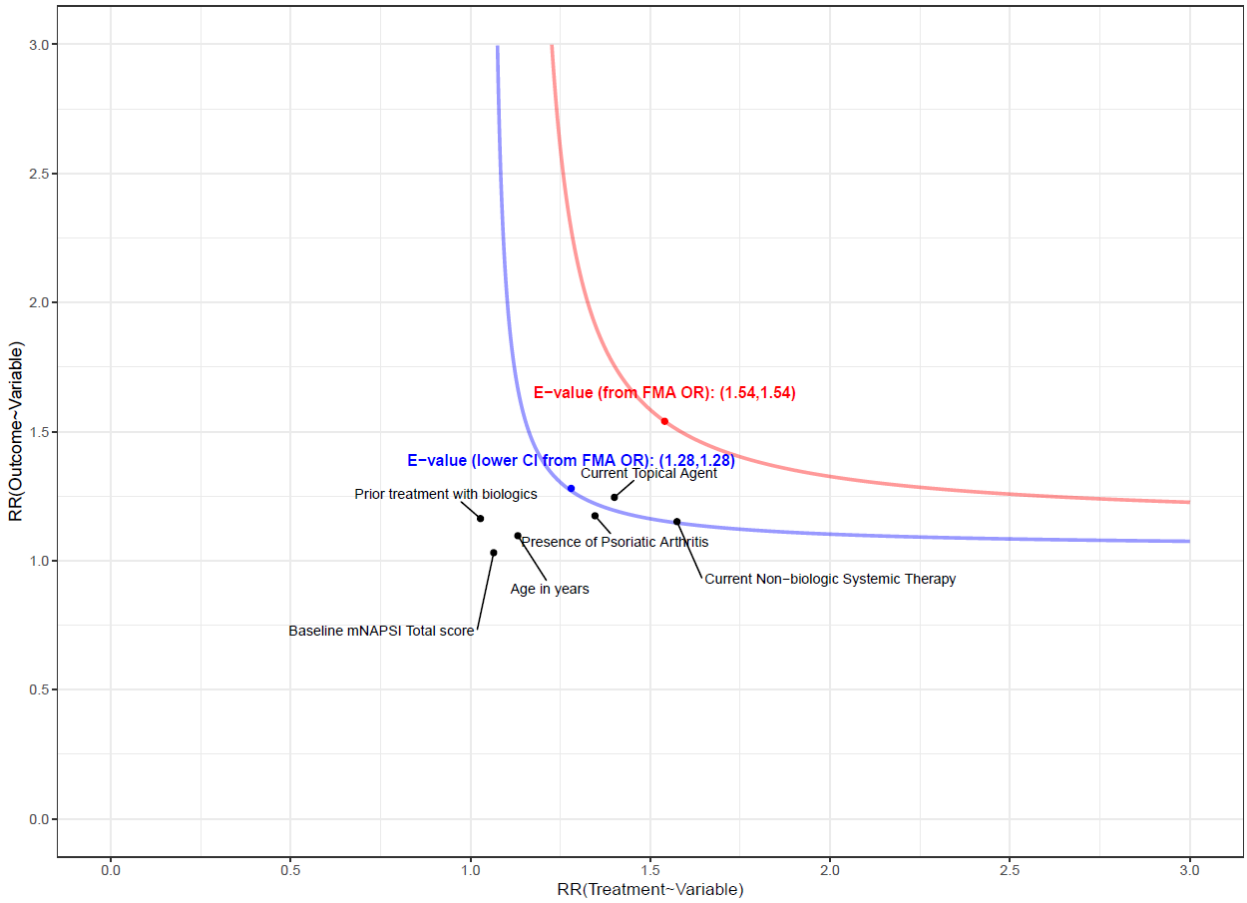

Figure 3: E-Value for the point estimate, for the lower confidence limit of the point estimate, and selected covariates for PASI 90 durability outcome at Month 12 (NRI), Anti-IL17A/RA vs Anti-IL-23*

*CI=Confidence interval, RR=Relative Risk, *Significant predictors of treatment selection. E-value is calculated from the FMA OR (converted to a RR) for the PASI 90 durability outcome (Anti-IL17A/RA vs Anti-IL-23), and a separate E-value (CI) is calculated from the lower confidence limit of the FMA OR (converted to a RR) for the PASI 90 durability outcome (Anti-IL17A/RA vs Anti-IL-23). The y-axis characterizes the strength of association between the unmeasured confounder and the outcome. The x-axis characterizes the extent to which the prevalence of the unmeasured confounder is unbalanced between the two treatment cohorts.*

References

1. Pinter A, Puig L, Schäkel K, Reich A, Zaheri S, Costanzo A, et al. Comparative effectiveness of biologics in clinical practice: week 12 primary outcomes from an international observational psoriasis study of health outcomes (PSoHO). Journal of the European Academy of Dermatology and Venereology. 2022;36(11):2087-100.
2. Pinter A, Costanzo A, Khattri S, Smith S, Carrascosa J, Tada Y, et al. Comparative Effectiveness and Durability of Biologics in Clinical Practice: Month 12 Outcomes from the International, Observational Psoriasis Study of Health Outcomes (PSoHO). Dermatology and Therapy. 2023:1-15.
3. McCaffrey DF, Griffin BA, Almirall D, Slaughter ME, Ramchand R, Burgette LF. A tutorial on propensity score estimation for multiple treatments using generalized boosted models. Statistics in medicine. 2013;32(19):3388-414.

4. Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B: Statistical Methodology. 2005;67(2):301-20.
5. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology. 1996;58(1):267-88.
6. Zagar A, Kadziola Z, Lipkovich I, Madigan D, Faries D. Evaluating bias control strategies in observational studies using frequentist model averaging. Journal of Biopharmaceutical Statistics. 2022;32(2):247-76.
7. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. Annals of internal medicine. 2017;167(4):268-74.