This is a repository copy of *Validating the assumptions of population adjustment:application of multilevel network meta-regression to a network of treatments for plaque psoriasis*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/189131/

Version: Accepted Version

Article:
Phillippo, David, Dias, Sofia orcid.org/0000-0002-2172-0221, Ades, A.E. et al. (5 more authors) (2023) Validating the assumptions of population adjustment:application of multilevel network meta-regression to a network of treatments for plaque psoriasis. Medical Decision Making. ISSN: 1552-681X

https://doi.org/10.1177/0272989X221117162

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Validating the assumptions of population adjustment: application of multilevel network meta-regression to a network of treatments for plaque psoriasis

David M. Phillippo, PhD[1]
University of Bristol, UK

Sofia Dias, PhD
University of York, UK and University of Bristol, UK

A. E. Ades, PhD
University of Bristol, UK

Mark Belger, BSc
Lilly UK, UK

Alan Brnabic, MA Stat
Eli Lilly Australia Pty. Limited, Australia

Daniel Saure, PhD and Yves Schymura, PhD
Lilly Deutschland GmbH, Germany

Nicky J. Welton, PhD
University of Bristol, UK

Running head: Validating assumptions of population adjustment

Word count: 5,442

Acknowledgements

The authors wish to thank three anonymous reviewers, whose insightful comments greatly improved this manuscript. This work was supported by the UK Medical Research Council, grant numbers MR/P015298/1, MR/R025223/1, and MR/W016648/1.

Declaration of Interests

DMP reports personal fees from UCB, Bristol Myers Squibb, and AstraZeneca outside of this work. SD, AEA, and NJW have no conflicts of interest to declare. MB, AB, DS, and YS are employees and shareholders of Eli Lilly and Company.

[1]University of Bristol, Canynge Hall, 39 Whatley Road, Bristol, BS8 2PS, UK. Email: david.phillippo@bristol.ac.uk

Draft social media content

How can the assumptions of population adjustment be assessed? @dmphillippo @sdias_stats @WeltonNicky and colleagues demonstrate, using multilevel network meta-regression to combine studies with individual patient data and aggregate data. #metaanalysis

Methods are applied to a network of treatments for plaque psoriasis with ordered categorical outcomes, implemented using the R package multinma https://dmphillippo.github.io/multinma/ #rstats

Validating the assumptions of population adjustment: application of multilevel network meta-regression to a network of treatments for plaque psoriasis

Abstract

Background: Network meta-analysis (NMA) and indirect comparisons combine aggregate data (AgD) from multiple studies on treatments of interest, but may give biased estimates if study populations differ. Population adjustment methods such as multilevel network meta-regression (ML-NMR) aim to reduce bias by adjusting for differences in study populations using individual patient data (IPD) from one or more studies under the conditional constancy assumption. A shared effect modifier assumption may also be necessary for identifiability. This paper aims to demonstrate how the assumptions made by ML-NMR can be assessed in practice to obtain reliable treatment effect estimates in a target population.
Methods: We apply ML-NMR to a network of evidence on treatments for plaque psoriasis with a mix of IPD AgD trials reporting ordered categorical outcomes. Relative treatment effects are estimated for each trial population and for three external target populations represented by a registry and two cohort studies. We examine residual heterogeneity and inconsistency, and relax the shared effect modifier assumption for each covariate in turn.
Results: Estimated population-average treatment effects were similar across study populations, as differences in the distributions of effect modifiers were small. Better fit was achieved with ML-NMR than NMA, and uncertainty was reduced by explaining within- and between-study variation. We found little evidence that the conditional constancy or shared effect modifier assumptions were invalid.
Conclusions: ML-NMR extends the NMA framework and addresses issues with previous population adjustment approaches. It coherently synthesises evidence from IPD and AgD studies in networks of any size whilst avoiding aggregation bias and non-collapsibility bias, allows for key assumptions to be assessed or relaxed, and can produce estimates relevant to a target population for decision-making.

Highlights

- Multilevel network meta-regression (ML-NMR) extends the network meta-analysis framework to synthesise evidence from networks of studies providing individual patient data or aggregate data whilst adjusting for differences in effect modifiers between studies (population adjustment). We apply ML-NMR to a network of treatments for plaque psoriasis with ordered categorical outcomes.

- We demonstrate for the first time how ML-NMR allows key assumptions to be assessed. We check for violations of conditional constancy of relative effects (such as unobserved effect modifiers) through residual heterogeneity and inconsistency, and the shared effect modifier assumption by relaxing this for each covariate in turn.

- Crucially for decision-making, population-adjusted treatment effects can be produced in any relevant target population. We produce population-average estimates for three external target populations, represented by the PsoBest registry and the PROSPECT and Chiricozzi 2019 cohort studies.

Introduction

Healthcare decision-making requires reliable estimates of relative treatment effects based, ideally, on high-quality randomised controlled trials (RCTs) comparing the treatments of interest, in a relevant target population. However, head-to-head RCTs between all relevant treatments are seldom available. Instead, comparisons are conducted using standard indirect comparison or network meta-analysis (NMA) methods using published aggregate data (AgD) from each study [1–4]. These methods assume that any variables that interact with treatment (effect modifiers) are balanced between study populations and that the study populations are representative of the target population of interest, which may not always be the case. Recently, population adjustment methods have been proposed that allow this assumption to be relaxed by adjusting for differences in effect modifiers using available individual patient data (IPD) from one or more studies [5–14].

We use the motivating example of a network of six active treatments plus placebo for moderate-to-severe plaque psoriasis from a previous systematic review [15], shown in Figure 1. IPD on outcomes and baseline covariates are available from four studies, and AgD consisting of summary outcomes and baseline covariate summaries are available for the remaining five studies (Table A.1). Outcomes of interest include success/failure to achieve at least 75%, 90%, or 100% improvement on the psoriasis area and severity index (PASI) scale at 12 weeks compared to baseline, denoted PASI 75, PASI 90, and PASI 100 respectively.

Matching-adjusted indirect comparison (MAIC) [8, 10] and simulated treatment comparison (STC) [9, 10] use reweighting or regression adjustment, respectively, to adjust the results of an IPD study to the population of an AgD study and estimate a population-adjusted indirect comparison. However, these methods were designed for two-study indirect comparisons (one IPD and one AgD). Although extensions to indirect comparisons with more than one IPD or AgD study have been proposed [16], these methods do not generalise easily to larger networks of studies and treatments which are frequently encountered in practice [17]. Warren et al. [15] previously analysed this network (Figure 1) using multiple MAIC analyses (following [16]), comparing ixekizumab Q2W against secukinumab 300 mg via either etanercept or placebo, and ustekinumab against secukinumab 300 mg via etanercept. However, these separate MAIC analyses are not coherent, in the same way that performing multiple pairwise meta-analysis does not ensure coherent estimates unlike those produced from a NMA [18]. Furthermore,

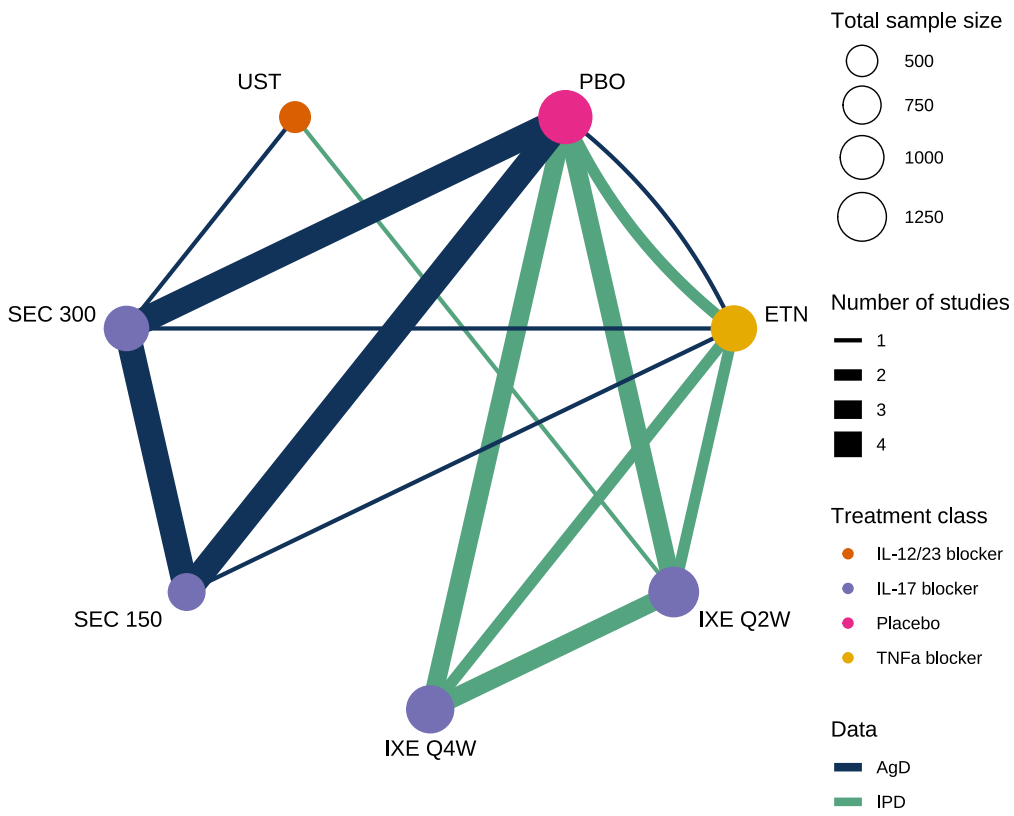

Figure 1: Network of studies comparing treatments for moderate-to-severe plaque psoriasis. PBO = placebo, IXE = ixekizumab, SEC = secukinumab, ETN = etanercept, UST = ustekinumab. IXE and SEC were each investigated with two different dosing regimens.

these multiple analyses are not independent as they reuse the same data, and none of them use all the available data (Table A.1). Crucially, unless further assumptions are made, MAIC and STC are limited to producing estimates for the population of the AgD study which may not match the target population for a decision [11, 12]. The AgD studies included in each of these analyses also differ, meaning that each analysis produces results for different target populations and thus are not directly compatible.

In the wider meta-analysis literature, several authors have considered combining both IPD and AgD in network meta-regression models to support the estimation of effect modifier interactions and increase statistical power and precision [5–7]. However, these methods simply "plug in" mean covariate values for the aggregate studies which results in aggregation bias in non-linear models [19]. Approaches that avoid this bias have been developed [20], but are only applicable when all covariates are binary.

Multilevel network meta-regression (ML-NMR) has recently been proposed to address many of the limitations with previous approaches [13, 14]. ML-NMR works by defining an

individual-level regression model, which is fitted directly to the individuals in the IPD studies, and incorporates summary outcomes from the AgD studies by integrating this individual-level model over the covariate distribution in each AgD study. As a result, ML-NMR appropriately synthesises networks of IPD and AgD studies of any size, adjusting for differences in effect modifiers whilst avoiding aggregation bias and non-collapsibility bias. Crucially for decision-making, ML-NMR can produce estimates of quantities of interest in any chosen target population, such as population-average treatment effects or absolute event probabilities in the decision target population. Furthermore, given sufficient data, ML-NMR may allow key assumptions to be assessed or relaxed [13]. ML-NMR is an extension of the standard NMA framework, reducing to standard AgD NMA when no covariates are adjusted for, and to full-IPD network meta-regression (the "gold standard" approach) when IPD are available from every study [13].

The key assumption made by all population adjustment methods in connected networks is that there are no unobserved effect modifiers, so that the relative effects are constant given the effect modifiers adjusted for—the conditional constancy of relative effects assumption [11, 12]. In sufficiently large networks this assumption may be assessed by checking for residual heterogeneity and inconsistency [4, 13, 21]. ML-NMR has previously been applied to a subset of the plaque psoriasis network consisting of only four studies, considering only the PASI 75 outcome [13]. However, no tests for heterogeneity or inconsistency were conducted. Furthermore, a shared effect modifier assumption [11] between ixekizumab and secukinumab (meaning that the effect modifier interaction parameters were assumed common for these treatments) was required to identify the model. Lastly, whilst higher PASI outcomes are more clinically meaningful, low numbers of observed events for PASI 90 and PASI 100 pose difficulties for estimation in stand-alone analyses of each PASI outcome.

In this paper, we demonstrate how the assumptions made by ML-NMR can be assessed in practice to obtain reliable treatment effect estimates in a target population. We jointly analyse the three PASI outcomes in an ordered categorical model, allowing information to be shared between outcomes and aiding estimation for the higher PASI outcomes. We show how to assess key assumptions using ML-NMR, including the conditional constancy of relative effects and shared effect modifier assumptions, which are untestable assumptions when using methods such as MAIC and STC. We produce estimates of population-average treatment effects and response probabilities for target populations represented by each of the studies in the network and for three external target populations.

Example: plaque psoriasis

A previous systematic review found nine studies comparing ixekizumab (IXE) every two weeks (Q2W) or four weeks (Q4W), secukinumab (SEC) at 150 mg dose or 300 mg dose, ustekinumab (UST) at a weight-based dose, and etanercept (ETN), along with placebo (PBO) [15]. These studies form the network in Figure 1 (summarised in Table A.1). IPD consisting of individual outcomes and baseline covariates were available from four studies, and AgD consisting of summary outcomes and baseline covariate summaries were extracted from published study reports on the remaining five studies (Table A.1). Table A.2 summarises the baseline covariate distributions in each of the nine studies. We jointly analyse the PASI 75, PASI 90, and PASI 100 outcomes at 12 weeks. We consider adjustments for duration of psoriasis, previous systemic treatment, body surface area covered, weight, and psoriatic arthritis, which were considered potential effect modifiers based on expert clinical opinion in previous analyses [13, 22]. In a decision-making context such as health technology assessment, we typically require principled selection of effect modifiers prior to analysis, either through expert opinion, systematic review, or quantitative analyses of external evidence [11]. Trials are typically underpowered to detect treatment-covariate interactions, and so selection based on post hoc criteria such as model fit, estimated effect size, or uncertainty, may result in bias due to over-interpretation of chance findings or omission of true effect modifiers. Sensitivity analyses may be performed, for example to compare the best fitting model to the pre-specified analysis model.

To be relevant for decision-making, estimates must be produced for a decision target population. Typically, the decision target population is not best represented by any of the RCTs in the network; instead, such a target population may be chosen based on expert clinical knowledge, or may be represented by a suitable registry or cohort study [11]. To illustrate, we produce population-adjusted estimates for three external target populations (Table A.3) [23–26].

Methods

Phillippo et al. [13] introduce the general ML-NMR model, which we extend here for ordered categorical outcomes following the approach taken for AgD NMA [4, 27].

ML-NMR for ordered categorical outcomes

At the individual level, we have ordered categorical outcomes: $< 75\%$ reduction in PASI score, $\geq 75\%$ and $< 90\%$ reduction, $\geq 90\%$ and $< 100\%$ reduction, and 100% reduction. These data are modelled using an ordered categorical likelihood, with the linear predictor for individual i in study j receiving treatment k given by

$$\eta_{jk}(\boldsymbol{x}_{ijk}) = \mu_j + \boldsymbol{x}_{ijk}^{\mathsf{T}}(\boldsymbol{\beta}_1 + \boldsymbol{\beta}_{2,k}) + \gamma_k, \tag{1}$$

which includes study-specific intercepts μ_j, individual-level covariates $\boldsymbol{x}_{ijk}$ with main effects $\boldsymbol{\beta}_1$ and treatment-covariate interactions $\boldsymbol{\beta}_{2,k}$ (corresponding to prognostic and effect modifying terms respectively), and individual-level treatment effects γ_k (full details in Appendix A.1.1). Consistency is assumed for both the individual-level treatment effects and the interactions [13]. We use the probit link function here for comparability with previous analyses, but the logit would be another suitable choice.

At the aggregate level, outcomes are vectors of summary outcome counts in each category. These summary data are given an ordered Multinomial likelihood, with average event probabilities in each category obtained by integrating the individual-level model over the covariate joint distribution in each arm of each AgD study. We calculate these integrals numerically using Quasi-Monte Carlo integration [13] with 1000 integration points. To perform this integration we require the covariate joint distributions in each of the AgD studies, however these are often not directly available: typically only marginal covariate information (e.g. means and standard deviations for continuous covariates, proportions for discrete covariates) is available from the AgD studies. We discuss how the covariate joint distributions can be reconstructed under some additional assumptions in the following section.

ML-NMR is typically implemented within a Bayesian framework. In this analysis, we place vague $\mathrm{N}(0, 10^2)$ prior distributions on each of the model parameters, except for the latent ordered cutpoints which are given improper uniform prior distributions $\mathrm{U}(-\infty, +\infty)$.

Using published marginal covariate information

To derive the aggregate-level model for ML-NMR through integration, the covariate joint distribution in each study is required. However, this is typically not available; instead, we have published marginal summary statistics for each covariate. We therefore reconstruct the

full joint distribution in each AgD study by making assumptions regarding the forms of the marginal distributions and the correlation structure [13]. Simulation studies have shown that these assumptions, even when incorrect, may have very little impact on the results in practice [28].

In the plaque psoriasis analysis, marginal distributions for each covariate in the AgD studies are chosen to match the reported summary statistics, based on theoretical properties and the observed distributions in the IPD studies: weight and duration are given a Gamma distribution to account for skewness, and body surface area as a percentage is given a scaled logit-Normal distribution. Previous systemic treatment and psoriatic arthritis are binary covariates. The correlation matrix for the covariates in the AgD studies is set to equal the weighted average of the correlation matrices in the IPD studies.

Checking model assumptions

The key assumption made by all population adjustment methods in connected networks is that there are no unobserved effect modifiers, so that the relative effects are constant given the effect modifiers adjusted for—the conditional constancy of relative effects assumption [11, 12]. This assumption implies that there is no remaining heterogeneity or inconsistency after adjusting for the given effect modifiers. In pairwise indirect comparisons this is an untestable assumption; however, in larger networks it is possible to assess this assumption by checking for residual heterogeneity and inconsistency [13], with techniques from standard AgD NMA [4, 21].

Assessing residual heterogeneity

We assess residual heterogeneity using a random effects ML-NMR model, where the linear predictor $\eta_{jk}(\boldsymbol{x})$ in equation (1) is replaced by

$$\eta_{jk}(\boldsymbol{x}) = \mu_j + \boldsymbol{x}^{\mathsf{T}}(\boldsymbol{\beta}_1 + \boldsymbol{\beta}_{2,k}) + \delta_{jk} \tag{2a}$$

$$\delta_{jk} \sim \mathrm{N}(\gamma_k, \tau^2) \tag{2b}$$

$$\mathrm{cor}(\delta_{ja}, \delta_{jb}) = 0.5 \tag{2c}$$

where $\delta_{j1} = 0$. The study-specific relative effects δ_{jk} within each study are multivariate Normal, with 0.5 correlations between non-treatment 1 arms under the assumption of common heterogeneity variance τ^2 [2]. Two-arm studies against treatment 1 have a single univariate

Normal random effect on the non-treatment 1 arm. We place a weakly-informative half-Normal prior distribution half-N$(0, 2.5^2)$ on the heterogeneity standard deviation τ, which means that there is 95% prior probability that τ lies between 0 and 5; $\tau = 5$ corresponds to 95% of true probit differences varying by ±10 between each study. Evidence of residual heterogeneity is assessed by comparing model fit under the fixed and random effects models using the Deviance Information Criterion (DIC) [29], and examining the posterior distribution of τ.

Assessing residual inconsistency

We assess residual inconsistency using an unrelated mean effects (UME) ML-NMR model [13, 21], where the linear predictor $\eta_{jk}(\boldsymbol{x})$ in equation (1) is replaced by

$$\eta_{jk}(\boldsymbol{x}) = \mu_j^{(t_{j1})} + \boldsymbol{x}^{\mathsf{T}}(\boldsymbol{\beta}_1 + \boldsymbol{\beta}_{2,k}) + \gamma_{t_{j1}k}, \tag{3}$$

where t_{j1} is the treatment in arm 1 of study j and we set $\gamma_{kk} = 0$ for all k. We place independent vague N$(0, 10^2)$ prior distributions on each of the γ_{ab} parameters, which now represent independent and unrelated treatment contrasts (instead of imposing the consistency equations). Under the UME model, the study baselines $\mu_j^{(t_{j1})}$ are now with respect to the treatments t_{j1} in arm 1 of each study, rather than the network reference treatment 1. A random effects UME model can also be fitted, replacing $\gamma_{t_{j1}k}$ with $\delta_{jt_{j1}k}$ in equation (3) which have a multivariate Normal structure analogous to that in equation (2) [13]. Evidence of residual inconsistency is assessed by comparing overall model fit under the consistency model and the UME inconsistency model using the DIC, and comparing the residual deviance contributions from each data point under either model. If random effects models are fitted the heterogeneity standard deviation τ should also be compared, because a reduction in estimated heterogeneity under the inconsistency model can indicate the presence of inconsistency [21]. As Donegan et al. [30] describe, inconsistency can also be present in the effect modifier interactions $\boldsymbol{\beta}_{2,k}$, which may be assessed through a similar approach to the UME model by placing independent prior distributions on the interactions on each contrast $\boldsymbol{\beta}_{2,ab}$. However, there are insufficient data to assess inconsistency in the interaction terms in this network.

Assessing the shared effect modifier assumption

Estimation of effect modifier interaction terms is data-intensive, requiring IPD or sufficiently many AgD studies with different covariate distributions on each treatment [13]. In the plaque psoriasis network we have only five AgD studies that include secukinumab arms, which is not sufficient to simultaneously estimate independent interactions for every effect-modifying covariate. Instead, we make the shared effect modifier assumption [11, 12] for ixekizumab and secukinumab, which both act as interleukin-17 blockers, meaning that these treatments share interaction terms and data requirements are shared across the class of treatments. That is, we set $\boldsymbol{\beta}_{2,k} = \boldsymbol{\beta}_{2,\mathrm{IL-17}}$ for each treatment k in this class. We assess this assumption directly, one covariate at a time, by splitting the common class interaction parameter for the covariate in question into independent treatment-specific interactions, whilst maintaining the common class interaction for the other covariates. Mathematically, for one covariate l^* in turn we let $\beta_{2,k;l^*}$ be independent for each treatment k, whilst the remaining covariates l retain the common class coefficients $\beta_{2,k;l} = \beta_{2,\mathrm{IL-17};l}$ for each treatment k. We then compare the posterior distributions of the interaction estimates from the independent interaction model to the corresponding common interaction estimate from the model making the shared effect modifier assumption, and compare overall model fit using the DIC. If residual heterogeneity or inconsistency are detected, then these may be assessed again for the independent interaction models to determine whether an invalid shared effect modifier assumption was contributing to heterogeneity or inconsistency.

Producing population-average estimates for populations of interest

Population-average estimates of quantities of interest to decision-making, such as average treatment effects and average event probabilities, can be produced by averaging estimates of individual-level quantities over the covariate joint distribution in the target population (see details in Appendix A.1.2) [13]. For decision-making based on cost-effectiveness models, the typical inputs are the population-average event probabilities for a cohort-based model (e.g. a decision tree or Markov model) or individual event probabilities in the population for an individual-based model (e.g. a discrete event simulation). The target population need not be one of the studies in the network; indeed, it is more likely that it is best represented by a registry or cohort study [11].

We produce estimates of population-adjusted average treatment effects in the three external

target populations represented by the PsoBest registry [23, 24], and the PROSPECT [25] and Chiricozzi 2019 [26] cohort studies. Following the process described in Appendix A.1.2, we require only the covariate summaries reported in Table A.3. Chiricozzi 2019 is the most different to the RCT populations, in patient age, bodyweight, disease severity, duration, and body surface area involvement, but there is still substantial overlap with the RCTs in the network so extrapolation is limited. In the absence of available data on covariate correlations in the external target populations, we use the weighted average correlation matrix computed from the IPD studies.

We also produce population-adjusted average event probabilities in the external target populations. In addition to the covariate summaries, this requires information on the response rates on one treatment (active or placebo) in the target population, to inform the baseline risk. PASI 75 response counts on secukinumab 300 mg at 12 weeks are available from PROSPECT and Chiricozzi 2019, from which we construct Beta distributions for the average PASI 75 event probability on secukinumab 300 mg in each target population, $\text{Beta}(1156, 1509 - 1165)$ and $\text{Beta}(243, 330 - 243)$ respectively, that serve as a reference against which the remaining population-average event probabilities on each treatment in each of these external target populations are produced (following Appendices A.1.2 and A.1.3). No information on event rates was available from PsoBest.

Statistical software

All analyses were carried out in R version 4.0.2 [31], using the package `multinma` [32]. Models are estimated in a Bayesian framework using Stan [33].

We fit the ML-NMR models described above, and for comparison we also fit standard AgD NMA models with no covariate adjustment. For all models, we assess convergence using the potential scale reduction factor $\widehat{R}$ for each parameter ensuring that $\widehat{R} < 1.01$ [34], and check that there are no divergent transitions [35]. Analysis code and data (including simulated IPD) are available in the online supplementary material, and a tutorial-style walkthrough of the analysis is available as a vignette in the `multinma` R package [32].

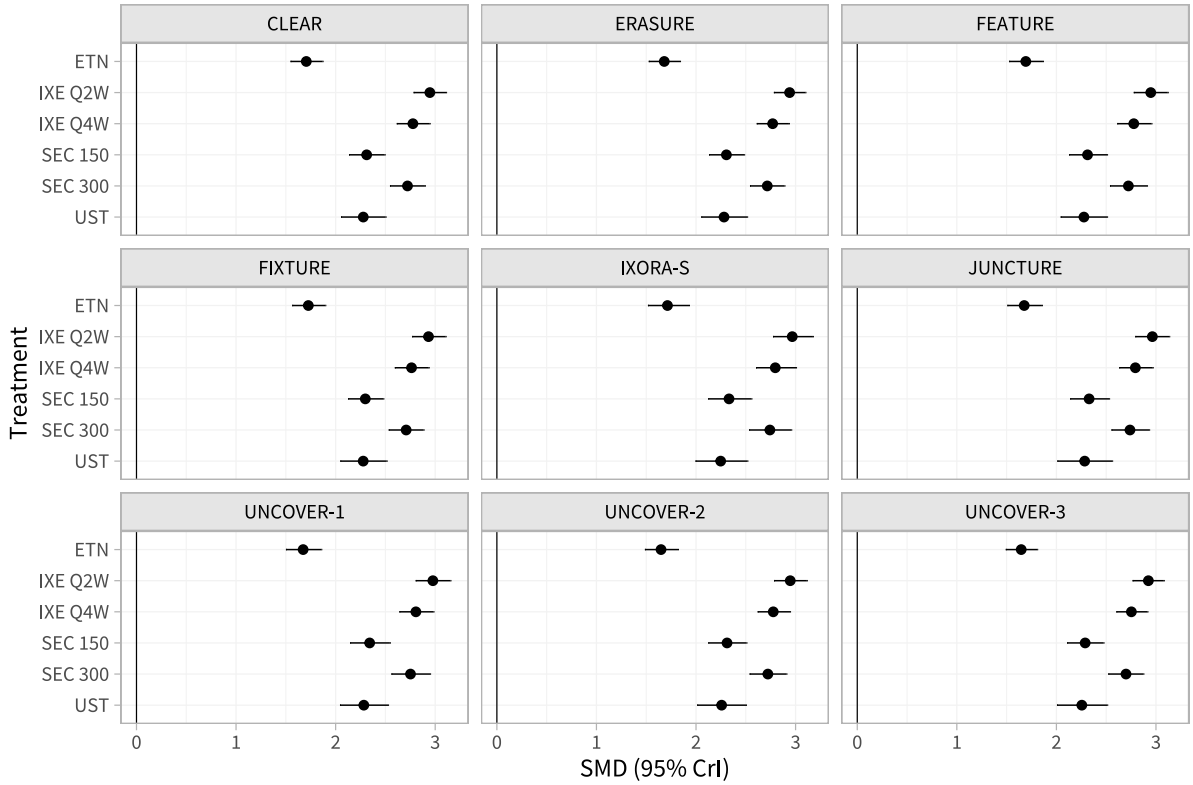

Figure 2: Estimated population-average treatment effects (standardised mean differences) for each treatment vs. placebo in the populations represented by each study in the network.

Results

ML-NMR model

Figure 2 and Table A.4 show the estimated population-average treatment effects for each treatment compared to placebo in each study population. Since the probit link function was used, these can be interpreted as standardised mean differences (SMDs) on the PASI scale. There is little variation in the population-average treatment effects between populations; this is due to the differences in effect modifier distributions between study populations (shown in Table A.2) being small when combined with the estimated strength of the interactions (Table 1).

The estimated proportion of individuals in each study population achieving each PASI outcome are shown in Figure 3, and listed in Tables A.5 to A.7.

We carried out sensitivity analyses removing covariates from the model (backward selection). The best-fitting model based on DIC was a model omitting body surface area (DIC 8810.1, residual deviance 8777.1). The estimated population-average treatment effects and event probabilities under this model were very similar to the pre-specified analysis model.

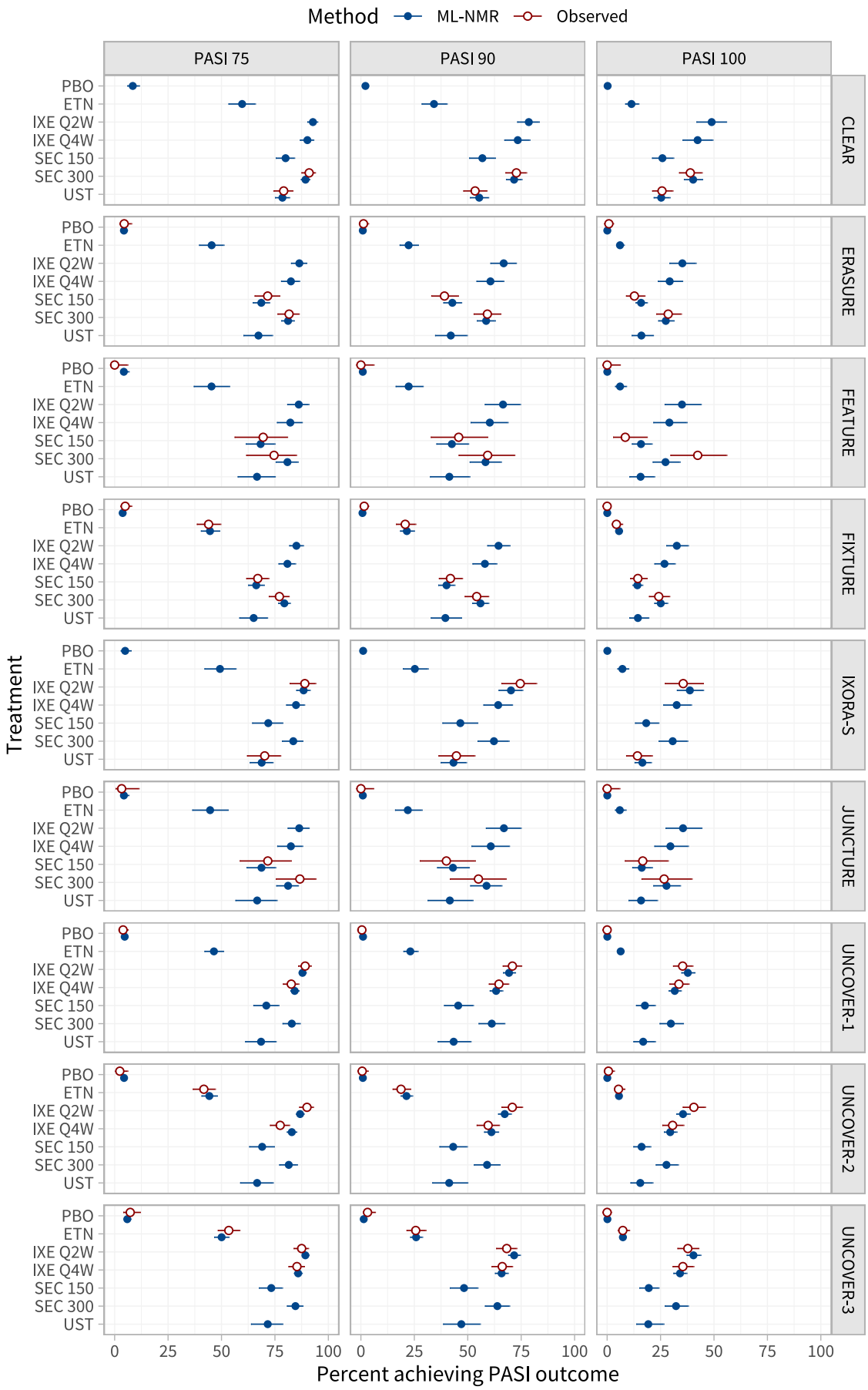

Figure 3: Estimated percentage of individuals achieving each PASI endpoint on each treatment, in the populations represented by each study in the network. For interpretability, these are given as inclusive probabilities (e.g. the probability of achieving 75% reduction or greater in PASI score). The observed event proportions are calculated from the event counts and sample sizes in each arm.

Table 1: Estimated interactions for each treatment class and potential effect modifier, and estimated individual-level treatment effects for an individual at the reference/centering values of the covariates (18.2 years disease duration, 29.8% body surface area, 89.3 kg weight, no previous systemic treatment or psoriatic arthritis), using the ML-NMR model. The shared effect modifier assumption was made for ixekizumab and secukinumab, which are both interleukin-17 blockers, and so these treatments share interaction terms. Etanercept and ustekinumab were treated as belonging to separate treatment classes (TNFα blocker and interleukin-12/23 blocker respectively) and given independent interaction terms. All estimates are standardised mean differences versus placebo, with 95% Credible Intervals.

	Treatment class		
	TNFα blocker	IL-17 blocker	IL-12/23 blocker
Effect modifier interaction			
Duration of psoriasis, per 10 years	0.17 (0.02, 0.32)	0.17 (0.02, 0.30)	0.12 (−0.08, 0.33)
Previous systemic use	0.11 (−0.27, 0.48)	0.13 (−0.21, 0.46)	−0.01 (−0.69, 0.67)
Body surface area, per 10%	0.04 (−0.06, 0.15)	0.01 (−0.09, 0.11)	0.05 (−0.09, 0.20)
Weight, per 10 kg	−0.09 (−0.17, −0.02)	−0.05 (−0.12, 0.02)	−0.04 (−0.14, 0.07)
Psoriatic arthritis	0.01 (−0.42, 0.49)	0.28 (−0.12, 0.71)	0.32 (−0.33, 1.02)
Reference individual treatment effect			
ETN	1.61 (1.35, 1.87)		
IXE Q2W		2.80 (2.55, 3.06)	
IXE Q4W		2.63 (2.38, 2.90)	
SEC 150		2.17 (1.91, 2.43)	
SEC 300		2.58 (2.33, 2.84)	
UST			2.21 (1.66, 2.76)

Assessing residual heterogeneity and inconsistency

Checking for residual heterogeneity using a RE ML-NMR model, the estimated heterogeneity standard deviation is 0.09 (0.01, 0.23), which is small compared to the relative treatment effects. Model fit statistics under each model are presented in Table 2. The DIC for the RE ML-NMR model is very similar to the FE ML-NMR model (8815.0 and 8814.9, respectively), and we would choose the more parsimonious fixed effect model based on DIC alone. Checking for residual inconsistency using an UME ML-NMR model (with fixed effects), this has a DIC of 8817.2 which is also very similar to the FE ML-NMR model assuming consistency. Plotting the residual deviance contributions from the consistency and inconsistency models (Figure A.1) does not suggest that any points fit better under either model. We conclude that there is no evidence for substantial residual heterogeneity or inconsistency after population adjustment: we have detected no failures in the assumptions for the FE ML-NMR model.

Results from fitting standard AgD NMA models (Appendix A.4) show little evidence for between-study heterogeneity or inconsistency, yet despite this the ML-NMR model still has a much lower DIC (Table 2). The ML-NMR model allows us to explain both between and within

Table 2: Model fit statistics for each ML-NMR and NMA model considered (FE, RE, and UME). p_D is a measure of the effective number of parameters. Residual deviance on 12,384 data points. Estimates and 95% Credible Intervals for the heterogeneity standard deviation τ are also presented for the RE models.

	ML-NMR			NMA		
	FE	RE	UME	FE	RE	UME
Residual deviance	8778.3	8773.2	8779.4	8931.4	8925.0	8932.3
p_D	36.3	41.8	37.8	16.8	22.4	17.7
DIC	8814.9	8815.0	8817.2	8948.2	8947.5	8950.1
τ	-	0.09 (0.01, 0.23)	-	-	0.09 (0.01, 0.24)	-

study variation, resulting in better fit and reduced uncertainty in contrast estimates across the study populations (Table A.4).

Assessing the shared effect modifier assumption

We assess the shared effect modifier assumption, one covariate at a time, by splitting the common interaction parameter for the covariate in question. Figure A.3 compares the posterior distributions of the independent interactions under each of the split models with the corresponding common interaction from the model making the shared effect modifier assumption for all covariates. In general, the independent interaction estimates are very similar to the common interaction estimates. The only exception is for weight, where there is some suggestion that this covariate may interact differently with the secukinumab treatment regimens to the ixekizumab regimens. However, the credible intervals for the secukinumab interactions are wide and overlap those for the ixekizumab regimens and the common interaction. In general, all of the independent interaction estimates are much more uncertain for the secukinumab regimens than for the ixekizumab regimens, as the secukinumab parameters are based solely on aggregate data, and the ixekizumab data largely drive the interaction estimates in the common interaction model. The DIC values (Table A.10) are higher for each of the independent interaction models than the common interaction model, except for the independent weight interaction model where the DIC values are nearly identical. Overall, there is some weak evidence that the shared effect modifier assumption (for the class of interleukin-17 blockers) may be invalid for weight; we consider this further in the discussion. Results from the ML-NMR model with independent weight interactions are given in Appendix A.5, and are very similar to those for the model making the shared effect modifier assumption for all covariates.

Producing population-average estimates for external target populations of interest

Estimated population-adjusted average treatment effects in the three external target populations are shown in Figure 4 and Table A.8. These are similar between the external target populations and the RCTs in the network, again because the differences in effect modifier distributions (Tables A.2 and A.3) are small when combined with the estimated strength of the interactions (Table 1). However, the estimated population-average probabilities of achieving each PASI outcome in the external target populations (Figure 5 and Table A.9) are lower than in the RCTs in the network, since the observed proportions achieving PASI 75 on secukinumab 300 mg are lower in the external target populations.

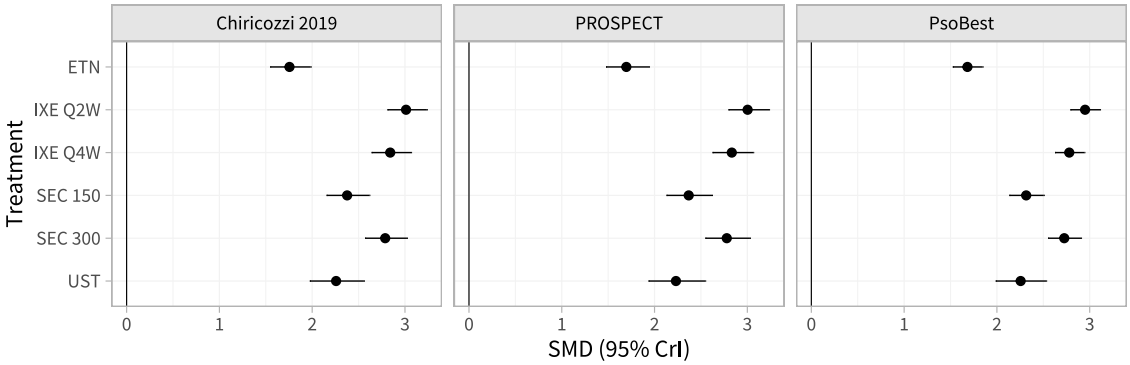

Figure 4: Estimated population-average treatment effects (standardised mean differences) for each treatment vs. placebo in each external target population.

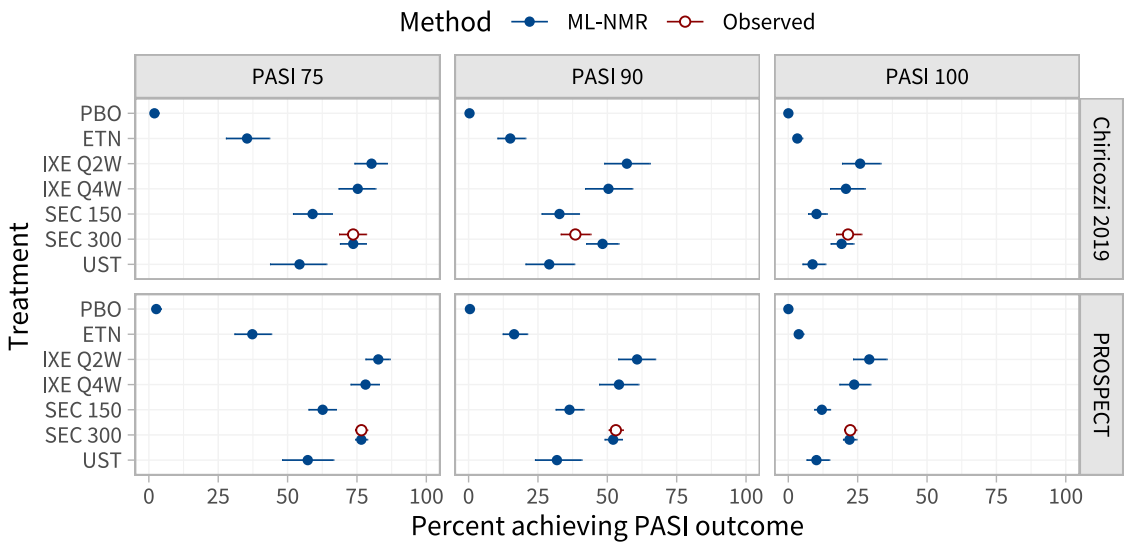

Figure 5: Estimated percentage of individuals achieving each PASI endpoint on each treatment, in each external target population with information on response rates. For interpretability, these are given as inclusive probabilities (e.g. the probability of achieving 75% reduction or greater in PASI score).

Discussion

In this paper, we have demonstrated an application of ML-NMR to synthesise PASI 75, 90, and 100 ordered categorical outcomes, whilst adjusting for variables considered to be potential effect modifiers, from a network containing a mixture of IPD and AgD studies. The ML-NMR analysis presented here also addresses several issues with previous MAIC analyses of this network [15]. In particular, ML-NMR makes full use of all available data in one coherent analysis that appropriately quantifies uncertainty (data are not reused), and estimated average treatment effects and average response probabilities for each PASI outcome can be produced in any of the included study populations or in an external target population, whichever is most relevant for decision-making. Moreover, this analysis produces more precise estimates than the previous MAIC analyses, and the 95% Credible Interval for the ixekizumab Q2W vs. secukinumab 300 relative effect (the focal comparison of the MAIC analyses) lies further from the null.

Synthesis of the three PASI outcomes together does require additional assumptions. These are not central to the ML-NMR analyses presented here, and are described elsewhere [4].

All population adjustment methods, including ML-NMR, assume that all effect modifiers have been suitably adjusted for so that the conditional constancy of relative effects assumption holds [11]. For pairwise indirect comparisons, including MAIC and STC, this is an untestable assumption. However, in larger networks, failures of conditional constancy of relative effects may manifest as residual heterogeneity or inconsistency, which can be assessed using standard techniques from AgD NMA. In this analysis, we used a random effects model and a (fixed effect) unrelated mean effects model to check for residual heterogeneity and inconsistency respectively, and concluded that there was no evidence of residual heterogeneity or inconsistency. However, just as in standard AgD NMA, these techniques may have low power, and detection of any unobserved effect modifiers relies on these being distributed differently at the aggregate level between studies. Moreover, these methods are fundamentally "in-sample" checks for heterogeneity and inconsistency—there may well be other omitted or unobserved effect modifiers that differ between the studies in the network and an external target population. In a decision-making context, this possibility should be considered when appraising the representativeness of the studies to the decision target population. Whilst we assessed inconsistency using the UME model, other inconsistency models such as node-splitting [13, 21] or design-by-treatment interactions [36] may also be implemented within the ML-NMR framework. Currently,

the `multinma` R package implements both UME and node-splitting models for assessing inconsistency [32].

In this analysis, we relied on the shared effect modifier assumption within the class of interleukin-17 blockers to support the estimation of effect modifier interaction terms. We assessed this assumption, one covariate at a time, by fitting independent interactions for the covariate in question. Exchangeable interactions within each treatment class could also be considered [13]; however, this was not possible given the data available. Relaxing the shared effect modifier assumption for multiple covariates at once is also a possibility, if sufficient data are available. However, for the purposes of checking this assumption relaxing one covariate at a time should be sufficient to highlight any violations. There was some weak evidence that the shared effect modifier assumption might not be valid for weight, as there were some differences between the independent interaction estimates and the common interaction estimate. However, there was high uncertainty in the independent interaction estimates for the secukinumab regimens, since these were estimated entirely from the five AgD studies, and model fit was unchanged. It is likely that this approach to assessing the shared effect modifier assumption has low power, particularly when data are lacking. There is also the possibility that these are chance findings due to fitting multiple models. Nevertheless, the analyses that we propose allow the validity and impact of this assumption to be investigated, even with these caveats, which was not previously possible. In general, differences in estimated interactions may be due to genuinely different interaction effects invalidating the shared effect modifier assumption, or due to an imbalance in unobserved effect modifiers across the studies that are correlated with weight. The latter leads to study-level confounding introducing bias into the interaction estimates ("ecological" bias) [37], and is the reason why meta-analytic studies of effect modification typically split interaction estimates into between-study interaction terms (which are susceptible to study-level confounding) and within-study interaction terms (which are then expected to be unbiased) [38, 39]. However, for the purposes of population adjustment we do not fit a split interaction model, since we must assume that there are no unobserved effect modifiers in order to produce population-adjusted estimates for a target population of interest. Nevertheless, it is possible to fit ML-NMR models where the interaction terms are split in this manner (given sufficient data), and this is an interesting avenue for further research.

When it is used, the shared effect modifier assumption is likely in many cases to be the most challenging to assess, since it is typically used to identify the model when data are insufficient.

However, even in cases where the shared effect modifier assumption is untestable, ML-NMR still retains other notable benefits over MAIC and STC [13, 28]. In particular, unlike MAIC, ML-NMR does not require full overlap between populations for stable estimation and remains unbiased under the usual extrapolation assumptions. Unlike STC, ML-NMR appropriately handles non-collapsible effect measures and does not combine incompatible (marginal and conditional) estimates. Moreover, the assumptions required by ML-NMR are still weaker than those for a standard NMA and indirect comparison, even when they are untestable. In an indirect comparison where the shared effect modifier assumption does not hold, the estimates from ML-NMR will only be unbiased and applicable within the AgD study population, as is the case with MAIC and STC. Small networks with multiple treatment classes and limited IPD are problematic for all current population adjustment methods, as there is likely to be insufficient information to combine the estimates in a coherent manner. In this analysis, even with a relatively small network (nine studies) we had sufficient data to relax and assess the key assumptions, checking for residual heterogeneity and inconsistency and assessing the shared effect modifier assumption.

When fitting network meta-regression models, as well as assuming consistency of the relative treatment effects there is a similar assumption of consistency on the effect modifier interactions. Donegan et al. [30] describe methods for assessing potential inconsistency in the effect modifier interactions, analogous to the node-splitting approach [21] for assessing inconsistency in treatment effects (described for ML-NMR models by Phillippo [14]). However, this network has only one potential loop where inconsistency could be present (via ustekinumab) since all other comparisons are made in multi-arm trials, and within this loop the data are insufficient to relax consistency and estimate unrelated interactions.

For the purposes of decision-making when effect modification is present, it is crucial that population-adjusted estimates are produced that are specific to the decision target population, whichever population-adjustment method is used [11]. Ideally, the decision target population is represented by a suitable registry or cohort study [11]; we have produced estimates relevant to the populations represented by registries and a cohort study. Each of these external target populations lay within the range of covariate values in the RCTs; when this is not the case, the validity of extrapolation should be considered. General guidance on when and how to use population-adjustment methods in a health technology assessment context is available [11].

When decisions are based on cost-effectiveness, the relevant inputs to the economic model are

(for a binary or categorical outcome) typically estimated event probabilities on each treatment in the target population. This requires information on the event probability on one treatment in the target population, in addition to the covariate distribution. If this information is not available, expert opinion or event probabilities from a similar population may be used.

When non-collapsible effect measures are used, for example odds ratios, hazard ratios, or the probit scale used here, ML-NMR typically targets a different population-average treatment effect estimand to MAIC (and indeed to standard AgD NMA based on event counts). The population-average treatment effects typically produced by ML-NMR target a population-average conditional estimand, the same as that targeted by a randomised controlled trial in the target population using an analysis adjusting for baseline prognostic factors. On the other hand, MAIC (and NMA of event counts) targets a (population-average) marginal estimand, the same as that targeted by a randomised controlled trial in the target population using an analysis without adjustment for baseline prognostic factors (or better from an adjusted analysis that has been marginalised over the covariates [40]). ML-NMR can also produce estimates of these marginal treatment effects, as we describe in Appendix A.1.2. Importantly, the marginal and conditional estimands have different interpretations when the effect measure is non-collapsible, and correspond to different decision questions. The population-average conditional treatment effect is the average effect between randomly selected treated and untreated individuals with the same covariates averaged over the distribution of covariates in the population [41], and answers the decision question "Which treatment has the greatest effect, on average, in this population?" The marginal population-average treatment effect is the average effect between randomly selected treated and untreated individuals in the population, regardless of their covariates [41], and answers the decision question "Which treatment minimises (or maximises) the average event probability in this population?" In the absence of effect modification these decision questions are aligned and always result in the same ranking of treatments, but this is not necessarily the case when effect modification is present, because then different treatments may be more effective for different individuals or subgroups within the population. Moreover, the population-average conditional treatment effects depend only on the distribution of effect modifying covariates, whereas the marginal treatment effects depend on the distributions of baseline risk, prognostic and effect modifying covariates, and on PASI cut-point. It is widely understood that when there is patient heterogeneity (including effect modification) health economic analyses need to appropriately handle this by averaging net benefit over the population

[42], for example by discrete event simulation [43, 44]. ML-NMR can produce estimates of both the conditional and marginal estimands, and necessary quantities for decision models such as average event probabilities or subgroup/individual event probabilities (see Appendix A.1.2).

Multilevel network meta-regression extends the standard NMA framework to incorporate IPD and AgD studies in networks of any size, adjusting for differences in effect modifiers between studies. It reduces to the "gold standard" IPD network meta-regression when IPD are available from every study, and reduces to standard AgD NMA when no covariates are adjusted for [13]. Moreover, we have demonstrated how techniques from the NMA literature can be used to assess the underlying assumptions of ML-NMR models. ML-NMR also addresses issues with previous methods such as MAIC, STC, and "plug in" meta-regression, since it synthesises networks of any size whilst avoiding aggregation bias and non-collapsibility bias, and can produce estimates of quantities of interest in any chosen target population. The R package `multinma` facilitates the application of ML-NMR models, making these methods available to a wider range of users [32].

Bibliography

1. Bucher HC, Guyatt GH, Griffith LE, and Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. Journal of Clinical Epidemiology. 1997; 50(6):683–691. DOI: 10.1016/s0895-4356(97)00049-8

2. Higgins JPT and Whitehead A. Borrowing strength from external trials in a meta-analysis. Statistics in Medicine. 1996; 15(24):2733–2749. DOI: 10.1002/(sici)1097-0258(19961230)15:24<2733::aid-sim562>3.0.co;2-0

3. Lu GB and Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Statistics in Medicine. 2004; 23(20):3105–3124. DOI: 10.1002/sim.1875

4. Dias S, Welton NJ, Sutton AJ, and Ades AE. NICE DSU Technical Support Document 2: A generalised linear modelling framework for pair-wise and network meta-analysis of randomised controlled trials. Tech. rep. National Institute for Health and Care Excellence, 2011. Available from: http://www.nicedsu.org.uk

5. Sutton AJ, Kendrick D, and Coupland CAC. Meta-analysis of individual- and aggregate-level data. Statistics in Medicine. 2008; 27(5):651–669. DOI: 10.1002/sim.2916

6. Saramago P, Sutton AJ, Cooper NJ, and Manca A. Mixed treatment comparisons using aggregate and individual participant level data. Statistics in Medicine. 2012; 31(28):3516–3536. DOI: 10.1002/sim.5442

7. Donegan S, Williamson P, D'Alessandro U, Garner P, and Tudur Smith C. Combining individual patient data and aggregate data in mixed treatment comparison meta-analysis: Individual patient data may be beneficial if only for a subset of trials. Statistics in Medicine. 2013; 32(6):914–930. DOI: 10.1002/sim.5584

8. Signorovitch JE, Wu EQ, Yu AP, Gerrits CM, Kantor E, Bao YJ, Gupta SR, and Mulani PM. Comparative Effectiveness Without Head-to-Head Trials A Method for Matching-Adjusted Indirect Comparisons Applied to Psoriasis Treatment with Adalimumab or Etanercept. Pharmacoeconomics. 2010; 28(10):935–945. DOI: 10.2165/11538370-000000000-00000

9. Caro JJ and Ishak KJ. No Head-to-Head Trial? Simulate the Missing Arms. Pharmacoeconomics. 2010; 28(10):957–967

10. Ishak KJ, Proskorovsky I, and Benedict A. Simulation and Matching-Based Approaches for Indirect Comparison of Treatments. Pharmacoeconomics. 2015; 33(6):537–549. DOI: 10.1007/s40273-015-0271-1

11. Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, and Welton NJ. NICE DSU Technical Support Document 18: Methods for population-adjusted indirect comparisons in submission to NICE. Tech. rep. National Institute for Health and Care Excellence, 2016. Available from: http://www.nicedsu.org.uk

12. Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, and Welton NJ. Methods for Population-Adjusted Indirect Comparisons in Health Technology Appraisal. Medical Decision Making. 2018; 38(2):200–211. DOI: 10.1177/0272989x17725740

13. Phillippo DM, Dias S, Ades AE, Belger M, Brnabic A, Schacht A, Saure D, Kadziola Z, and Welton NJ. Multilevel Network Meta-Regression for population-adjusted treatment comparisons. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2020; 183(3):1189–1210. DOI: 10.1111/rssa.12579

14. Phillippo DM. Calibration of Treatment Effects in Network Meta-Analysis using Individual Patient Data. Available from https://research-information.bris.ac.uk/. PhD thesis. University of Bristol, 2019

15. Warren RB, Brnabic A, Saure D, Langley RG, See K, Wu JJ, Schacht A, Mallbris L, and Nast A. Matching-adjusted indirect comparison of efficacy in patients with moderate-to-severe plaque psoriasis treated with ixekizumab vs. secukinumab. British Journal of Dermatology. 2018; 178(5):1064–1071. DOI: 10.1111/bjd.16140

16. Belger M, Brnabic A, Kadziola Z, Petto H, and Faries D. Inclusion of Multiple Studies in Matching Adjusted Indirect Comparisons (MAIC). *ISPOR 20th Annual International Meeting*. Philadelphia, PA, USA, 2015. DOI: 10.1016/j.jval.2015.03.196

17. Phillippo DM, Dias S, Elsada A, Ades AE, and Welton NJ. Population Adjustment Methods for Indirect Comparisons: A Review of National Institute for Health and Care Excellence Technology Appraisals. International Journal of Technology Assessment in Health Care. 2019; 35(03):221–228. DOI: 10.1017/S0266462319000333

18. Caldwell DM, Ades AE, and Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ. 2005; 331(7521):897–900. DOI: 10.1136/bmj.331.7521.897

19. Berlin JA, Santanna J, Schmid CH, Szczech LA, and Feldman HI. Individual patient-versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. Statistics in Medicine. 2002; 21(3):371–387. DOI: 10.1002/sim.1023

20. Jansen JP. Network meta-analysis of individual and aggregate level data. Research Synthesis Methods. 2012; 3(2):177–190. DOI: 10.1002/jrsm.1048

21. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, and Ades AE. NICE DSU Technical Support Document 4: Inconsistency in networks of evidence based on randomised controlled trials. Tech. rep. National Institute for Health and Care Excellence, 2011. Available from: http://www.nicedsu.org.uk

22. Strober B, Brnabic A, Schacht A, Mallbris L, See K, Warren RB, and Nast A. Indirect Comparison of Ixekizumab and Secukinumab Using Matched-Adjusted Indirect Comparisons. Oral presentation presented at 25th Congress of the European Academy of Dermatology and Venereology. Vienna, Austria, 2016

23. Reich K, Mrowietz U, Radtke MA, Thaci D, Rustenbach SJ, Spehr C, and Augustin M. Drug safety of systemic treatments for psoriasis: results from The German Psoriasis Registry

PsoBest. Archives of Dermatological Research. 2015; 307(10):875–883. DOI: 10.1007/s00403-015-1593-8

24. Augustin M et al. German psoriasis registry PsoBest: objectives, methodology and baseline data. JDDG: Journal der Deutschen Dermatologischen Gesellschaft. 2014; 12(1):48–57. DOI: 10.1111/ddg.12233

25. Thaçi D, Körber A, Kiedrowski R, Bachhuber T, Melzer N, Kasparek T, Duetting E, Kraehn-Senftleben G, Amon U, and Augustin M. Secukinumab is effective in treatment of moderate-to-severe plaque psoriasis: real-life effectiveness and safety from the PROSPECT study. Journal of the European Academy of Dermatology and Venereology. 2019; 34(2):310–318. DOI: 10.1111/jdv.15962

26. Chiricozzi A et al. Secukinumab demonstrates improvements in absolute and relative psoriasis area severity indices in moderate-to-severe plaque psoriasis: results from a European, multicentric, retrospective, real-world study. Journal of Dermatological Treatment. 2019; 31(5):476–483. DOI: 10.1080/09546634.2019.1671577

27. Woolacott N et al. Etanercept and efalizumab for the treatment of psoriasis: a systematic review. Health Technology Assessment. 2006; 10(46). DOI: 10.3310/hta10460

28. Phillippo DM, Dias S, Ades AE, and Welton NJ. Assessing the performance of population adjustment methods for anchored indirect comparisons: A simulation study. Statistics in Medicine. 2020. Available online 4th October 2020. DOI: 10.1002/sim.8759

29. Spiegelhalter DJ, Best NG, Carlin BP, and Linde A van der. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2002; 64(4):583–639. DOI: 10.1111/1467-9868.00353

30. Donegan S, Welton NJ, Tudur Smith C, D'Alessandro U, and Dias S. Network meta-analysis including treatment by covariate interactions: Consistency can vary across covariate values. Research Synthesis Methods. 2017; 8(4):485–495. DOI: 10.1002/jrsm.1257

31. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria, 2020. Available from: https://www.R-project.org/

32. Phillippo DM. multinma: Network Meta-Analysis of Individual and Aggregate Data in Stan. R package. 2020. DOI: 10.5281/zenodo.3904454. Available from: https://cran.r-project.org/package=multinma

33. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, and Riddell A. Stan: A Probabilistic Programming Language. Journal of Statistical Software. 2017; 76(1). DOI: 10.18637/jss.v076.i01

34. Vehtari A, Gelman A, Simpson D, Carpenter B, and Bürkner PC. Rank-Normalization, Folding, and Localization: An Improved $\widehat{R}$ for Assessing Convergence of MCMC. Bayesian Analysis. 2020. DOI: 10.1214/20-ba1221

35. Betancourt MJ and Girolami M. Hamiltonian Monte Carlo for Hierarchical Models. 2013. arXiv: 1312.0906 [stat.ME]

36. Jackson D, Barrett JK, Rice S, White IR, and Higgins JP. A design-by-treatment interaction model for network meta-analysis with random inconsistency effects. Statistics in Medicine. 2014; 33(21):3639–3654. DOI: 10.1002/sim.6188

37. Greenland S. Divergent biases in ecologic and individual-level studies. Statistics in Medicine. 1992; 11(9):1209–1223. DOI: 10.1002/sim.4780110907

38. Hua H, Burke DL, Crowther MJ, Ensor J, Tudur Smith C, and Riley RD. One-stage individual participant data meta-analysis models: estimation of treatment-covariate interactions must avoid ecological bias by separating out within-trial and across-trial information. Statistics in Medicine. 2016; 36(5):772–789. DOI: 10.1002/sim.7171

39. Fisher DJ, Carpenter JR, Morris TP, Freeman SC, and Tierney JF. Meta-analytical methods to identify who benefits most from treatments: daft, deluded, or deft approach? BMJ. 2017 :j573. DOI: 10.1136/bmj.j573

40. Zhang Z. Estimating a Marginal Causal Odds Ratio Subject to Confounding. Communications in Statistics - Theory and Methods. 2008; 38(3):309–321. DOI: 10.1080/03610920802200076

41. Kahan BC, Jairath V, Doré CJ, and Morris TP. The risks and rewards of covariate adjustment in randomized trials: an assessment of 12 outcomes from 8 studies. Trials. 2014; 15(1). DOI: 10.1186/1745-6215-15-139

42. Welton NJ, Soares MO, Palmer S, Ades AE, Harrison D, Shankar-Hari M, and Rowan KM. Accounting for Heterogeneity in Relative Treatment Effects for Use in Cost-Effectiveness Models and Value-of-Information Analyses. Medical Decision Making. 2015; 35(5):608–621. DOI: 10.1177/0272989x15570113

43. Karnon J, Stahl J, Brennan A, Caro JJ, Mar J, and Möller J. Modeling using Discrete Event Simulation: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-4. Value in Health. 2012; 15(6):821–827. DOI: 10.1016/j.jval.2012.04.013

44. Karnon J and Afzali HHA. When to Use Discrete Event Simulation (DES) for the Economic Evaluation of Health Technologies? A Review and Critique of the Costs and Benefits of DES. PharmacoEconomics. 2014; 32(6):547–558. DOI: 10.1007/s40273-014-0147-9

A Appendix

A.1 Statistical methods

A.1.1 ML-NMR for ordered categorical outcomes

With ordered categorical outcomes in M categories, the IPD outcomes are given by $y_{ijk} \in \{1, \ldots, M\}$. For the plaque psoriasis network, there are $M = 4$ categories, and we let $y_{ijk} \in \{1, \ldots, 4\}$ correspond to less than 75% reduction in PASI score (i.e. failure to achieve PASI 75), $\geq$ 75% and $<$ 90% reduction (achieving PASI 75 but not PASI 90), $\geq$ 90% and $<$ 100% reduction (achieving PASI 90 but not PASI 100), and 100% reduction (achieving PASI 100), respectively, for an individual i in study j receiving treatment k. The individual-level model for ordered categorical outcomes is given by

$$y_{ijk} \sim \text{Categorical}(p_{ijk;1}, \ldots, p_{ijk;M}) \tag{A.1a}$$

$$p_{ijk;m} = \Phi(\eta_{jk}(\boldsymbol{x}_{ijk}) - c_{m-1}) - \Phi(\eta_{jk}(\boldsymbol{x}_{ijk}) - c_m) \quad \text{for } m = 1, \ldots, M \tag{A.1b}$$

$$\eta_{jk}(\boldsymbol{x}_{ijk}) = \mu_j + \boldsymbol{x}_{ijk}^{\mathsf{T}}(\boldsymbol{\beta}_1 + \boldsymbol{\beta}_{2,k}) + \gamma_k \tag{A.1c}$$

where the event probability in each category is $p_{ijk;m}$ and $\sum_{m=1}^{4} p_{ijk,m} = 1$ for each i, j, k. The category event probabilities are transformed onto the linear predictor scale using the probit link function $\Phi^{-1}(\cdot)$ (the Normal inverse cumulative distribution function). We use the probit link function here for comparability with previous analyses, but another suitable choice would be the logit link function. The linear predictor for an individual on treatment k in trial j with covariate vector $\boldsymbol{x}_{ijk}$ is $\eta_{jk}(\boldsymbol{x}_{ijk})$. The parameters μ_j are study-specific baselines, $\boldsymbol{\beta}_1$ are coefficients for prognostic variables, and $\boldsymbol{\beta}_{2,k}$ are coefficients for effect modifiers specific to each treatment k. The effect of treatment k (at the individual level), γ_k, is defined with respect to the network reference treatment 1, and we set $\gamma_1 = 0$ and $\boldsymbol{\beta}_{2,1} = \mathbf{0}$. Some coefficients in $\boldsymbol{\beta}_1$ or $\boldsymbol{\beta}_{2,k}$ may be set to zero, if it is known that a particular covariate is not prognostic or effect modifying respectively. The consistency equations on the individual-level treatment effects, where the relative effect of treatment b vs. a is $\gamma_{ab} = \gamma_b - \gamma_a$, are implicit in this model formulation, which is equivalent to the widely-used "baseline shift" formulation of NMA models [4]. A similar set of consistency equations are also implicit on the effect modifier interactions $\boldsymbol{\beta}_{2,ab} = \boldsymbol{\beta}_{2,b} - \boldsymbol{\beta}_{2,a}$. The latent

cutpoints between categories c_m are subject to the ordering constraints

$$c_0 < c_1 < \cdots < c_M, \tag{A.2}$$

and, we set $c_0 = -\infty$, $c_1 = 0$, and $c_M = +\infty$. For the plaque psoriasis analysis, the latent cutpoint c_1 corresponds to achieving PASI 75, c_2 corresponds to achieving PASI 90, and c_3 corresponds to achieving PASI 100. A very similar model has previously been proposed for standard AgD NMA of ordered count data [4, 27]; indeed, model (A.1) reduces to this when there are no covariates.

Priors for the latent cutpoints are most straightforward to specify on the differences between adjacent cutpoints, for example $c_m - c_{m-1} \sim \mathrm{U}(0, u_c)$ for $m = 2, \ldots, M-1$ with an appropriate upper bound u_c (as used by Dias et al. [4] with $u_c = 5$), so that the ordering constraints (A.2) are satisfied. When fitting the model in Stan, the ordering constraints (A.2) are guaranteed by declaring the c_m to be an `ordered` vector, so prior distributions can be placed directly on the cutpoints if desired. In this analysis, we place improper uniform prior distributions $\mathrm{U}(-\infty, +\infty)$ on c_2 and c_3, which are automatically truncated to satisfy the ordering constraints (A.2). We also place vague $\mathrm{N}(0, 10^2)$ prior distributions on each of the parameters μ_j, $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_{2,j}$, and γ_k.

Aggregate outcomes are vectors of summary outcome counts in each category $\boldsymbol{y}_{\bullet jk} = (y_{\bullet jk;1}, \ldots, y_{\bullet jk;M})^{\mathsf{T}}$. We work with these category counts in "exclusive" format, where individuals are only counted in the highest category they achieve (as opposed to "inclusive" counts where individuals are counted in every category up to and including the highest category achieved; it is a straightforward matter of addition or subtraction to convert between formats). These summary data are given a Multinomial likelihood, with the average event probabilities in each category obtained by integrating the individual-level model (A.1b) over the covariate joint distribution $f_{jk}(\cdot)$ in each arm of each AgD study:

$$\boldsymbol{y}_{\bullet jk} \sim \mathrm{Multi}(\bar{p}_{jk;1}, \ldots, \bar{p}_{jk;M}; N_{jk}) \tag{A.3a}$$

$$\bar{p}_{jk;m} = \bar{p}_{jk;m} = \bar{q}_{jk;m-1} - \bar{q}_{jk;m} \quad \text{for } m = 1, \ldots, M, \text{ where} \tag{A.3b}$$

$$\bar{q}_{jk;m} = \int_{\mathfrak{X}} \Phi(\eta_{jk}(\boldsymbol{x}) - c_m) f_{jk}(\boldsymbol{x})\, d\boldsymbol{x} \tag{A.3c}$$

where $\bar{p}_{jk;m}$ are the average event probabilities in each category and $N_{jk} = \sum_k y_{\bullet jk;m}$ is is the sample size in each arm. We compute the integrals for $\bar{q}_{jk;m}$ in (A.3c) using Quasi-Monte Carlo integration [13] with $\tilde{N} = 1000$ integration points $\tilde{\boldsymbol{x}}_{jk}$ drawn from joint distribution $f_{jk}(\cdot)$ of the

covariates on each treatment k in study j, so that

$$\bar{q}_{jk;m} \simeq \tilde{N}^{-1} \sum \Phi(\eta_{jk}(\tilde{\boldsymbol{x}}_{jk}) - c_m).$$

Accounting for studies reporting a subset of categories

It is not uncommon for some studies to report only a subset of categories. These studies can be incorporated at either the individual or aggregate level by modifying equations Equation (A.1) or Equation (A.3), respectively, to involve the relevant latent cutpoints [4]. That is, given a reported set of categories $s_1 < \cdots < s_{M_j}$ in study j, the individual-level event probabilities in (A.1b) for the reported categories become

$$p_{ijk;s_m} = \Phi(\eta_{jk}(\boldsymbol{x}_{ijk}) - c_{s_{m-1}}) - \Phi(\eta_{jk}(\boldsymbol{x}_{ijk}) - c_{s_m}) \quad \text{for } m = 1, \ldots, M_j.$$

Similarly at the aggregate level, the average event probabilities (A.3b) in each of the reported categories become

$$\bar{p}_{jk;s_m} = \bar{p}_{jk;m} = \bar{q}_{jk;s_{m-1}} - \bar{q}_{jk;s_m} \quad \text{for } m = 1, \ldots, M_j.$$

A.1.2 Producing population-average estimates for populations of interest

Population-average estimates of quantities of interest to decision-making, such as average treatment effects and average event probabilities, can be produced by averaging estimates of individual-level quantities over the covariate joint distribution in the target population [13]. For decision-making based on cost-effectiveness models, the typical inputs are the population-average event probabilities for a cohort-based model (e.g. a decision tree or Markov model) or individual event probabilities for an individual-based model (e.g. a discrete event simulation). The target population need not be one of the studies in the network; indeed, it is more likely represented by a registry or cohort study [11].

To estimate the proportion of individuals achieving each PASI endpoint in a given population, we first note that the probability of an individual in population P with covariate values $\boldsymbol{x}$ achieving PASI endpoint m or greater is

$$q_{k;m(P)}(\boldsymbol{x}) = \Phi\big(\mu_{(P)} + \boldsymbol{x}^{\mathsf{T}}(\boldsymbol{\beta}_1 + \boldsymbol{\beta}_{2,k}) + \gamma_k - c_m\big), \tag{A.4}$$

where $m = 1$ corresponds to PASI 75, $m = 2$ to PASI 90, and $m = 3$ to PASI 100; $\mu_{(P)}$ is the baseline probit probability of achieving PASI 75 in population P. A distribution for $\mu_{(P)}$ is required, which may be estimated from external data on the population P. Appendix A.1.3 describes techniques for deriving a distribution for $\mu_{(P)}$ in some common evidence scenarios for baseline response. Here, we produce estimates for each of the observed study populations and simply use the intercepts μ_j estimated in the model. We then average (A.4) over the covariate joint distribution $f_{(P)}(\cdot)$ in population P. This may be performed in general by

$$\bar{q}_{k;m(P)} = N^{-1} \sum_{i=1}^{N} q_{k;m(P)}(\boldsymbol{x}_{i(P)}), \tag{A.5}$$

where $\boldsymbol{x}_{i(P)}$ are N samples from $f_{(P)}(\cdot)$. For example, $\boldsymbol{x}_{i(P)}$ may be $N_{(P)}$ covariate values of individuals in a large representative sample from the population P, or these may be $\tilde{N}$ integration points if the joint covariate distribution in the population has been constructed from summary information or from a small sample of IPD (following section "Using published marginal covariate information").

For quantities on the linear predictor scale such as population-average treatment effects, averaging over the joint covariate distribution is equivalent to "plugging-in" mean covariate values from the population of interest. The population-average treatment effect between treatments a and b in population P is estimated using

$$d_{ab(P)} = \bar{\boldsymbol{x}}_{(P)}^{\mathsf{T}}(\boldsymbol{\beta}_{2,b} - \boldsymbol{\beta}_{2,a}) + \gamma_b - \gamma_a, \tag{A.6}$$

where $\bar{\boldsymbol{x}}_{(P)}$ is the vector of mean covariate values in population P.

The population-average treatment effects $d_{ab(P)}$ produced by equation (A.6) are population-average *conditional* treatment effects, the average effect between randomly selected treated and untreated individuals with the same covariates in the population. On the other hand, MAIC targets the (population-average) *marginal* treatment effect, the average effect between randomly selected treated and untreated individuals in the population, regardless of their covariates. ML-NMR can also produce estimates of these marginal treatment effects, as

$$\Delta_{ab;m(P)} = \Phi^{-1}\big(\bar{q}_{b;m(P)}\big) - \Phi^{-1}\big(\bar{q}_{a;m(P)}\big), \tag{A.7}$$

for any two treatments b and a in population P for outcome category m, where $\bar{q}_{k;m(P)}$ are

estimated using equation (A.5). Analogous approaches can be used to express marginal effects in terms of risk differences or risk ratios for interpretation. Due to non-collapsibility, $\Delta_{ab;m(P)}$ will differ from $d_{ab(P)}$, and—like the average category event probabilities $\bar{q}_{k;m(P)}$—will also differ by the distributions of baseline event probabilities and prognostic (as well as effect-modifying) variables between populations, and between categories. Moreover, rankings based on the marginal effects $\Delta_{ab;m(P)}$ (or equivalently on the average event probabilities $\bar{q}_{k;m(P)}$) may differ from those based on the population-average conditional effects $d_{ab(P)}$ when there is effect modification, because the estimands correspond to different decision questions. For decision-making based on cost-effectiveness models, the typical inputs are the population-average event probabilities $\bar{q}_{k;m(P)}$ for a cohort-based model (e.g. a decision tree or Markov model) or individual event probabilities $q_{k;m(P)}(\boldsymbol{x})$ for an individual-based model (e.g. a discrete event simulation).

A.1.3 Transforming information on baseline response

As described in Appendix A.1.2, to produce estimates of quantities of interest such as predicted probabilities of response we require a distribution for the baseline response probit probability $\mu_{(P)}$ in the target population P. For ML-NMR models the interpretation of $\mu_{(P)}$ is the probit probability of achieving the lowest response category (PASI 75) on the network reference treatment 1, for an individual in population P at the reference value of the covariates. However, we may not have information on baseline response in this format. Here we describe some simple transformations for obtaining the baseline response information in the required format.

For example, if we have baseline response information as a distribution for $\mu^{(k)}_{(P);m}$ the probit probability of response for an individual in P on treatment k in response category m at the reference level of the covariates, we can recover $\mu_{(P)}$ from the relation $\mu_{(P)} = \mu^{(k)}_{(P);m} - \gamma_k + c_m$. This calculation can be performed within the ML-NMR model code at the time of fitting, or afterwards using posterior samples of γ_k and c_m from the fitted model and a sample (of the same size) from the distribution of $\mu^{(k)}_{(P);m}$.

Moreover, it is perhaps more usual to have information on the baseline response probability at the population level, say the average response probability on treatment k in response category m in the population, $\bar{q}_{k;m(P)}$. We can obtain a sample of the distribution for $\mu_{(P)}$ from a sample of the distribution for $\bar{q}_{k;m(P)}$ by solving equation (A.5) for $\mu_{(P)}$. This can be performed using numerical optimisation, either within the ML-NMR model code at the time of fitting or using posterior samples from the model parameters and a sample of the same length for $\bar{q}_{k;m(P)}$, solving

(A.5) at every sample. By definition any link function (including the probit and logit) is smooth and monotonic, so we have a well-behaved optimisation problem with a unique solution.

A.2 Details of included studies and target populations

Table A.1: Structure of studies and treatments included in the network; • indicates treatment arms included in studies from which IPD were available, ∘ indicates treatment arms included in studies from which AgD were available.

	CLEAR	ERASURE	FEATURE	FIXTURE	IXORA-S	JUNCTURE	UNCOVER-1	UNCOVER-2	UNCOVER-3
Placebo	·	∘	∘	∘	·	∘	•	•	•
Etanercept	·	·	·	∘	·	·	·	•	•
Ixekizumab Q2W	·	·	·	·	•	·	•	•	•
Ixekizumab Q4W	·	·	·	·	·	·	•	•	•
Secukinumab 150 mg	·	∘	∘	∘	·	∘	·	·	·
Secukinumab 300 mg	∘	∘	∘	∘	·	∘	·	·	·
Ustekinumab	∘	·	·	·	•	·	·	·	·

Table A.2: Baseline covariate summaries for each study. Reported sample size for UNCOVER-2 and 3 after removing two individuals from each study with missing weight, and one individual from IXORA-S with missing body surface area. Statistics are mean (SD) unless otherwise specified. * Covariate considered a potential effect modifier, to be included in population adjustment.

	CLEAR (N = 676)	ERASURE (N = 738)	FEATURE (N = 177)	FIXTURE (N = 1306)	IXORA-S (N = 259)	JUNCTURE (N = 182)	UNCOVER-1 (N = 1296)	UNCOVER-2 (N = 1219)	UNCOVER-3 (N = 1339)
Age, years	44.9 (13.8)	45.1 (13.1)	45.9 (13.9)	44.5 (12.9)	43.7 (12.8)	44.7 (13.8)	45.7 (12.9)	45.0 (13.0)	45.7 (13.1)
* Body surface area, per cent	32.3 (17.3)	31.9 (18.1)	32.0 (17.3)	34.4 (18.9)	29.1 (17.0)	27.4 (14.7)	27.7 (17.3)	26.0 (16.5)	28.3 (17.1)
* Duration of psoriasis, years	17.8 (12.1)	17.4 (11.8)	19.5 (13.0)	16.5 (12.0)	18.0 (11.8)	20.5 (13.4)	19.6 (11.9)	18.7 (12.5)	18.2 (12.2)
Baseline PASI score	21.6 (8.3)	22.1 (9.4)	20.8 (8.2)	23.7 (10.2)	21.3 (8.5)	20.1 (7.3)	20.1 (8.0)	19.6 (7.2)	20.9 (8.2)
* Previous systemic treatment (%)	67.5	63.0	67.2	64.0	92.3	55.5	71.3	64.2	57.1
* Psoriatic arthritis (%)	18.2	23.2	16.4	14.7	14.7	23.1	26.3	23.6	20.5
Male (%)	71.2	69.0	66.1	71.1	67.2	68.7	68.1	67.0	68.2
* Weight, kg	87.3 (21.0)	88.5 (23.8)	91.6 (24.4)	83.3 (20.8)	87.7 (22.3)	91.6 (25.3)	92.2 (23.8)	91.6 (22.2)	91.2 (23.5)

Table A.3: Covariate summaries for each external target population. Statistics are mean (SD) unless otherwise specified.
* Covariate considered a potential effect modifier, to be included in population adjustment.

	Chiricozzi 2019 (N = 330)	PROSPECT (N = 1509)	PsoBest (N = 2556)
Age, years	51.9 (14.6)	48.1 (13.7)	47.4 (14.1)
Baseline PASI score	16.6 (8.2)	17.7 (12.5)	14.7 (9.7)
* Body surface area, per cent	23.0 (16.8)	18.7 (18.4)	24.0 (20.5)
* Duration of psoriasis, years	16.9 (10.8)	19.6 (13.5)	18.2 (14.1)
Male (%)	68.2	62.4	60.1
* Previous systemic treatment (%)	90.6	91.0	54.0
* Psoriatic arthritis (%)	21.5	20.2	20.7
* Weight, kg	78.3 (15.9)	87.5 (20.3)	85.0 (19.1)

A.3 Additional results figures and tables

Table A.4: Estimated population-average treatment effects (standardised mean differences) and 95% Credible Intervals for each treatment compared to placebo, plus selected focal comparisons of interest between licensed dose regimens, in each study population using the ML-NMR model and for the FE NMA.

	ML-NMR study population									
Contrast	CLEAR	ERASURE	FEATURE	FIXTURE	IXORA-S	JUNCTURE	UNCOVER-1	UNCOVER-2	UNCOVER-3	FE NMA
ETN vs. PBO	1.71 (1.55, 1.87)	1.68 (1.53, 1.84)	1.69 (1.53, 1.87)	1.73 (1.56, 1.90)	1.71 (1.52, 1.93)	1.68 (1.51, 1.85)	1.67 (1.50, 1.86)	1.65 (1.49, 1.82)	1.65 (1.49, 1.81)	1.61 (1.47, 1.74)
IXE Q2W vs. PBO	2.95 (2.78, 3.11)	2.94 (2.78, 3.10)	2.95 (2.78, 3.12)	2.93 (2.77, 3.11)	2.97 (2.77, 3.17)	2.96 (2.79, 3.13)	2.98 (2.80, 3.15)	2.95 (2.78, 3.11)	2.92 (2.77, 3.08)	2.86 (2.73, 3.00)
IXE Q4W vs. PBO	2.78 (2.61, 2.94)	2.77 (2.61, 2.93)	2.78 (2.61, 2.96)	2.76 (2.59, 2.94)	2.80 (2.60, 3.00)	2.79 (2.63, 2.97)	2.81 (2.64, 2.98)	2.78 (2.62, 2.94)	2.75 (2.60, 2.91)	2.68 (2.55, 2.83)
SEC 150 vs. PBO	2.31 (2.13, 2.49)	2.31 (2.13, 2.48)	2.31 (2.13, 2.51)	2.30 (2.13, 2.48)	2.33 (2.12, 2.56)	2.33 (2.14, 2.53)	2.34 (2.15, 2.55)	2.31 (2.12, 2.51)	2.29 (2.11, 2.47)	2.18 (2.02, 2.33)
SEC 300 vs. PBO	2.72 (2.55, 2.90)	2.72 (2.54, 2.89)	2.72 (2.54, 2.91)	2.71 (2.53, 2.88)	2.74 (2.53, 2.96)	2.74 (2.55, 2.93)	2.75 (2.56, 2.95)	2.72 (2.54, 2.91)	2.70 (2.52, 2.87)	2.58 (2.43, 2.73)
UST vs. PBO	2.28 (2.05, 2.50)	2.28 (2.05, 2.51)	2.28 (2.04, 2.51)	2.28 (2.05, 2.51)	2.25 (2.00, 2.52)	2.28 (2.01, 2.56)	2.28 (2.04, 2.53)	2.26 (2.01, 2.50)	2.26 (2.01, 2.51)	2.14 (1.95, 2.34)
SEC 300 vs. IXE Q2W	−0.22 (−0.39, −0.07)	−0.22 (−0.39, −0.07)	−0.22 (−0.39, −0.07)	−0.22 (−0.39, −0.07)	−0.22 (−0.39, −0.07)	−0.22 (−0.39, −0.07)	−0.22 (−0.39, −0.07)	−0.22 (−0.39, −0.07)	−0.22 (−0.39, −0.07)	−0.28 (−0.43, −0.14)
UST vs. IXE Q2W	−0.67 (−0.86, −0.47)	−0.66 (−0.87, −0.45)	−0.67 (−0.87, −0.47)	−0.66 (−0.86, −0.44)	−0.72 (−0.92, −0.52)	−0.68 (−0.92, −0.43)	−0.69 (−0.89, −0.49)	−0.69 (−0.90, −0.47)	−0.67 (−0.90, −0.43)	−0.72 (−0.90, −0.54)
UST vs. SEC 300	−0.44 (−0.60, −0.28)	−0.43 (−0.61, −0.26)	−0.45 (−0.61, −0.28)	−0.43 (−0.60, −0.26)	−0.49 (−0.69, −0.30)	−0.45 (−0.67, −0.24)	−0.47 (−0.64, −0.29)	−0.46 (−0.65, −0.28)	−0.44 (−0.64, −0.24)	−0.44 (−0.58, −0.29)

Table A.5: Estimated proportion of individuals achieving PASI 75 on each treatment in each study population, along with 95% Credible Intervals, using ML-NMR combining information from all PASI endpoints. For interpretability, these are given as inclusive probabilities (i.e. the probability of achieving 75% reduction or greater in PASI score).

	Treatment						
Study population	Placebo	Etanercept	Ixekizumab Q2W	Ixekizumab Q4W	Secukinumab 150 mg	Secukinumab 300 mg	Ustekinumab
CLEAR	8.47 (5.95, 11.53)	59.64 (53.03, 66.12)	92.34 (89.19, 94.81)	89.73 (85.76, 92.94)	79.60 (74.62, 84.04)	88.89 (86.17, 91.14)	78.84 (75.08, 82.45)
ERASURE	4.33 (3.06, 5.80)	45.33 (39.56, 51.02)	86.13 (82.12, 89.57)	82.25 (77.65, 86.39)	68.73 (64.94, 72.36)	80.99 (77.76, 84.04)	67.07 (59.89, 73.89)
FEATURE	4.29 (2.57, 6.60)	45.28 (37.14, 53.67)	85.62 (79.68, 90.57)	81.67 (74.92, 87.76)	68.00 (60.76, 74.79)	80.40 (74.64, 85.59)	66.74 (57.24, 75.41)
FIXTURE	3.74 (2.67, 5.06)	44.55 (40.21, 48.91)	84.83 (80.96, 88.15)	80.70 (76.17, 84.69)	66.53 (62.14, 70.64)	79.35 (75.89, 82.48)	64.64 (57.70, 71.31)
IXORA-S	4.95 (2.89, 7.64)	49.44 (41.85, 56.96)	87.22 (82.47, 91.10)	83.48 (77.67, 88.53)	70.22 (62.25, 77.65)	82.25 (76.80, 87.00)	70.62 (63.16, 77.16)
JUNCTURE	4.33 (2.57, 6.61)	44.66 (35.96, 53.02)	85.67 (79.63, 90.52)	81.74 (74.79, 87.62)	68.17 (60.70, 74.91)	80.47 (74.35, 85.73)	67.21 (57.73, 75.97)
UNCOVER-1	4.72 (3.49, 6.10)	46.42 (41.95, 50.89)	86.91 (84.11, 89.36)	83.16 (80.05, 86.03)	69.87 (63.96, 75.59)	81.85 (77.27, 86.11)	69.50 (61.95, 76.65)
UNCOVER-2	4.42 (3.22, 5.80)	44.31 (40.63, 47.88)	85.86 (83.27, 88.24)	81.91 (78.93, 84.64)	68.09 (62.21, 73.71)	80.53 (75.95, 84.80)	67.52 (59.62, 74.86)
UNCOVER-3	5.93 (4.39, 7.63)	50.08 (46.73, 53.43)	88.78 (86.48, 90.74)	85.40 (82.77, 87.69)	73.02 (67.56, 78.07)	84.20 (79.88, 87.80)	71.92 (64.72, 78.50)

Table A.6: Estimated proportion of individuals achieving PASI 90 on each treatment in each study population, along with 95% Credible Intervals, using ML-NMR combining information from all PASI endpoints. For interpretability, these are given as inclusive probabilities (i.e. the probability of achieving 90% reduction or greater in PASI score).

	Treatment						
Study population	Placebo	Etanercept	Ixekizumab Q2W	Ixekizumab Q4W	Secukinumab 150 mg	Secukinumab 300 mg	Ustekinumab
CLEAR	2.12 (1.33, 3.18)	34.21 (28.25, 40.56)	78.14 (72.34, 83.22)	73.06 (66.39, 79.06)	56.94 (50.47, 63.19)	71.46 (67.09, 75.38)	55.22 (50.31, 60.26)
ERASURE	0.89 (0.57, 1.30)	22.28 (18.09, 26.76)	66.93 (61.04, 72.61)	60.90 (54.56, 66.97)	43.58 (39.41, 47.83)	59.02 (54.65, 63.32)	41.19 (33.83, 48.92)
FEATURE	0.89 (0.46, 1.53)	22.33 (16.41, 29.05)	66.19 (57.53, 74.37)	60.14 (51.10, 69.24)	42.86 (35.44, 50.47)	58.26 (50.66, 65.80)	40.96 (31.41, 50.83)
FIXTURE	0.74 (0.47, 1.09)	21.46 (18.23, 24.87)	64.68 (58.93, 70.04)	58.49 (52.43, 64.39)	41.05 (36.23, 45.98)	56.56 (51.72, 61.19)	38.47 (31.51, 45.57)
IXORA-S	1.06 (0.52, 1.86)	25.38 (19.53, 31.81)	68.48 (60.76, 75.38)	62.47 (53.87, 70.27)	45.02 (36.32, 53.96)	60.59 (52.72, 67.93)	45.06 (36.97, 53.05)
JUNCTURE	0.90 (0.45, 1.56)	21.94 (15.77, 28.57)	66.37 (57.62, 74.46)	60.35 (51.06, 69.28)	43.13 (35.30, 50.86)	58.47 (50.43, 66.13)	41.54 (31.83, 51.58)
UNCOVER-1	1.01 (0.68, 1.41)	23.13 (19.79, 26.71)	68.16 (64.37, 71.78)	62.19 (58.39, 66.16)	44.89 (38.55, 51.55)	60.28 (54.13, 66.41)	43.90 (35.98, 52.26)
UNCOVER-2	0.92 (0.60, 1.32)	21.33 (18.63, 24.12)	66.33 (62.83, 69.87)	60.23 (56.40, 63.94)	42.80 (36.63, 49.12)	58.29 (52.15, 64.49)	41.61 (33.60, 50.02)
UNCOVER-3	1.34 (0.89, 1.88)	25.86 (23.12, 28.66)	71.41 (67.93, 74.50)	65.68 (61.91, 69.08)	48.57 (42.24, 54.84)	63.82 (57.61, 69.50)	46.67 (38.76, 54.91)

Table A.7: Estimated proportion of individuals achieving PASI 100 on each treatment in each study population, along with 95% Credible Intervals, using ML-NMR combining information from all PASI endpoints.

	Treatment						
Study population	Placebo	Etanercept	Ixekizumab Q2W	Ixekizumab Q4W	Secukinumab 150 mg	Secukinumab 300 mg	Ustekinumab
CLEAR	0.23 (0.12, 0.40)	11.40 (8.41, 14.97)	49.15 (41.88, 56.31)	42.73 (35.44, 50.13)	26.63 (21.43, 32.19)	40.78 (36.21, 45.46)	24.53 (20.72, 28.75)
ERASURE	0.08 (0.04, 0.13)	5.99 (4.30, 7.90)	36.09 (30.24, 42.08)	30.25 (24.70, 36.00)	16.93 (14.00, 20.09)	28.52 (24.52, 32.58)	14.98 (10.79, 19.85)
FEATURE	0.08 (0.03, 0.15)	6.05 (3.77, 9.00)	35.43 (27.29, 44.31)	29.65 (22.18, 38.24)	16.52 (11.94, 21.70)	27.93 (21.69, 34.75)	14.93 (9.66, 21.34)
FIXTURE	0.06 (0.03, 0.10)	5.56 (4.28, 7.03)	33.65 (28.16, 39.23)	27.96 (22.82, 33.41)	15.26 (12.21, 18.76)	26.30 (22.09, 30.84)	13.30 (9.66, 17.54)
IXORA-S	0.09 (0.04, 0.19)	7.20 (4.76, 10.18)	37.35 (29.72, 45.30)	31.37 (23.92, 39.42)	17.67 (12.30, 24.12)	29.61 (23.08, 36.81)	17.31 (12.42, 22.81)
JUNCTURE	0.08 (0.03, 0.15)	5.93 (3.60, 8.82)	35.71 (27.15, 44.62)	29.93 (22.14, 38.40)	16.76 (11.79, 22.24)	28.22 (21.58, 35.30)	15.32 (9.94, 21.93)
UNCOVER-1	0.09 (0.05, 0.14)	6.32 (4.96, 7.88)	37.29 (33.96, 40.95)	31.35 (28.31, 34.81)	17.80 (13.78, 22.43)	29.66 (24.47, 35.33)	16.67 (11.90, 22.30)
UNCOVER-2	0.08 (0.04, 0.13)	5.52 (4.48, 6.69)	35.28 (31.91, 38.94)	29.45 (26.12, 32.81)	16.39 (12.51, 20.74)	27.81 (22.64, 33.32)	15.18 (10.62, 20.56)
UNCOVER-3	0.13 (0.07, 0.20)	7.40 (6.13, 8.77)	40.85 (37.17, 44.53)	34.70 (31.15, 38.47)	20.29 (15.81, 25.34)	32.93 (27.22, 38.97)	18.44 (13.41, 24.32)

Table A.8: Estimated population-average treatment effects (standardised mean differences) and 95% Credible Intervals for each treatment compared to placebo, plus selected focal comparisons of interest between licensed dose regimens, in each external target population using ML-NMR.

	External target population		
Contrast	PsoBest	PROSPECT	Chiricozzi 2019
ETN vs. PBO	1.68 (1.52, 1.85)	1.70 (1.48, 1.94)	1.76 (1.55, 1.98)
IXE Q2W vs. PBO	2.95 (2.79, 3.11)	3.00 (2.80, 3.23)	3.01 (2.81, 3.24)
IXE Q4W vs. PBO	2.78 (2.63, 2.94)	2.83 (2.63, 3.06)	2.84 (2.64, 3.07)
SEC 150 vs. PBO	2.32 (2.13, 2.51)	2.37 (2.13, 2.62)	2.38 (2.16, 2.62)
SEC 300 vs. PBO	2.73 (2.55, 2.91)	2.78 (2.55, 3.03)	2.79 (2.57, 3.02)
UST vs. PBO	2.26 (1.99, 2.53)	2.23 (1.94, 2.55)	2.26 (1.97, 2.56)
SEC 300 vs. IXE Q2W	−0.22 (−0.39, −0.07)	−0.22 (−0.39, −0.07)	−0.22 (−0.39, −0.07)
UST vs. IXE Q2W	−0.69 (−0.94, −0.43)	−0.77 (−1.01, −0.54)	−0.75 (−0.99, −0.52)
UST vs. SEC 300	−0.47 (−0.69, −0.25)	−0.55 (−0.78, −0.31)	−0.53 (−0.75, −0.31)

Table A.9: Estimated proportion of individuals achieving each PASI outcome on each treatment in each external target population with information on response rates, along with 95% Credible Intervals, using the ML-NMR model combining information from all PASI endpoints. For interpretability, these are given as inclusive probabilities (e.g. the probability of achieving 75% reduction or greater in PASI score).

	Treatment						
Population	Placebo	Etanercept	Ixekizumab Q2W	Ixekizumab Q4W	Secukinumab 150 mg	Secukinumab 300 mg	Ustekinumab
Chiricozzi 2019							
PASI 75	2.51 (1.10, 4.79)	38.73 (28.68, 49.38)	80.08 (73.84, 85.33)	75.17 (68.25, 81.55)	59.23 (52.08, 66.16)	73.62 (68.82, 78.27)	60.68 (48.06, 72.83)
PASI 90	0.45 (0.16, 1.00)	17.25 (10.86, 24.93)	57.25 (48.90, 65.28)	50.73 (42.26, 59.18)	33.41 (26.80, 40.43)	48.73 (42.79, 54.90)	34.58 (23.35, 47.35)
PASI 100	0.03 (0.01, 0.08)	4.00 (2.03, 6.82)	26.55 (19.90, 33.65)	21.46 (15.59, 28.25)	10.79 (7.52, 14.74)	19.96 (15.81, 24.55)	11.24 (5.93, 18.56)
PROSPECT							
PASI 75	3.45 (1.70, 6.10)	41.49 (32.86, 50.75)	82.56 (78.11, 86.47)	78.03 (72.74, 82.74)	62.93 (58.08, 67.77)	76.60 (74.47, 78.70)	65.13 (54.15, 75.41)
PASI 90	0.67 (0.27, 1.38)	19.33 (13.39, 26.21)	61.00 (54.40, 67.03)	54.62 (47.63, 61.35)	37.14 (32.15, 42.61)	52.65 (49.40, 55.79)	39.12 (28.33, 50.49)
PASI 100	0.05 (0.02, 0.13)	4.80 (2.76, 7.50)	29.98 (24.25, 35.99)	24.57 (19.17, 30.44)	12.87 (9.94, 16.30)	22.96 (20.18, 25.94)	13.76 (8.18, 20.93)

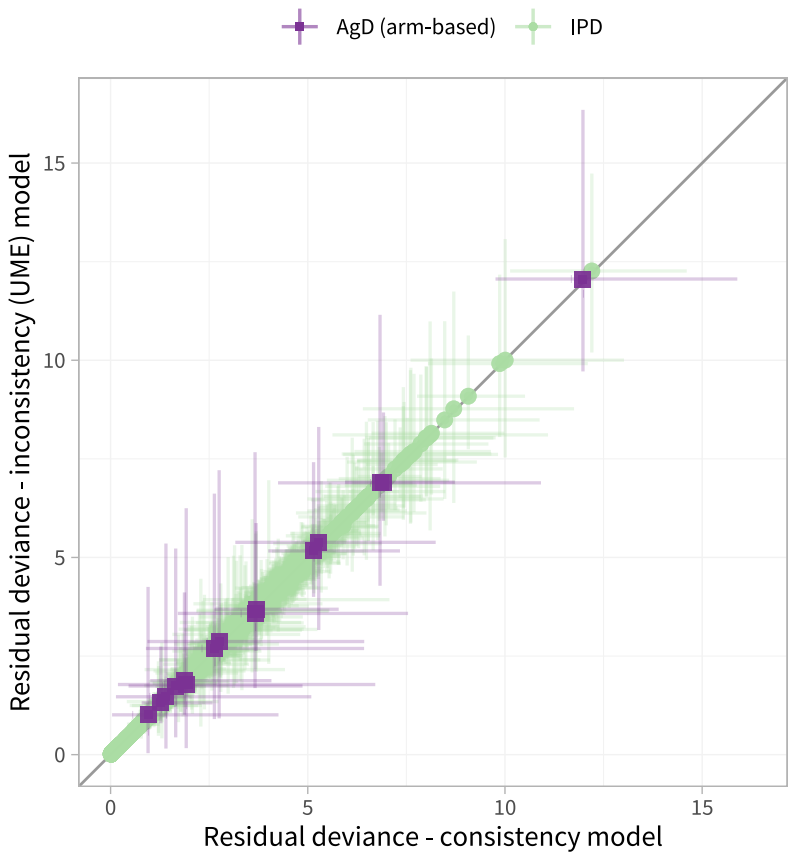

Figure A.1: Residual deviance contributions (posterior mean and 95% Credible Interval) under a fixed effect ML-NMR model assuming consistency, and under a fixed effect ML-NMR inconsistency (UME) model. Here, all of the posterior mean residual deviance contributions lie on the line of equality; all points fit equally well under either model.

A.4 Comparison with aggregate data NMA

For models to be comparable using DIC, we need the input data to be the same. The DIC values in Table 2 for the NMAs are therefore calculated considering the residual deviance of each IPD point separately (i.e. with the uncollapsed data, not data collapsed into summary counts as is usual), so that direct comparison can be made with the ML-NMR models. The collapsed and uncollapsed NMA models are mathematically equivalent, and the conclusions from model comparison between the NMA models are also the same since the residual deviance and DIC are equivalent up to a fixed constant.

The estimated heterogeneity standard deviation from the RE NMA without covariate adjustment was 0.09 (0.01, 0.24), which is small compared to the magnitude of the relative effects (Table A.4). The DIC values for the FE and RE NMA models without covariate adjustment were 8948.2 and 8947.5 respectively (see Table 2); there is little difference between these models, and we would choose the more parsimonious fixed effect model based on DIC alone. The DIC for an UME model which does not make the consistency assumption is 8950.1 which is a little higher than the FE consistency model, and a plot of the residual deviance contributions under each model (Figure A.2) shows that all data points are fit equally well under either model, altogether suggesting no evidence of inconsistency. However, despite a lack of evidence for between-study heterogeneity or inconsistency, the ML-NMR model has a much lower DIC of 8814.9. The ML-NMR model allows us to explain both between and within study variation, resulting in better fit and reduced uncertainty in contrast estimates across the study populations (Table A.4).

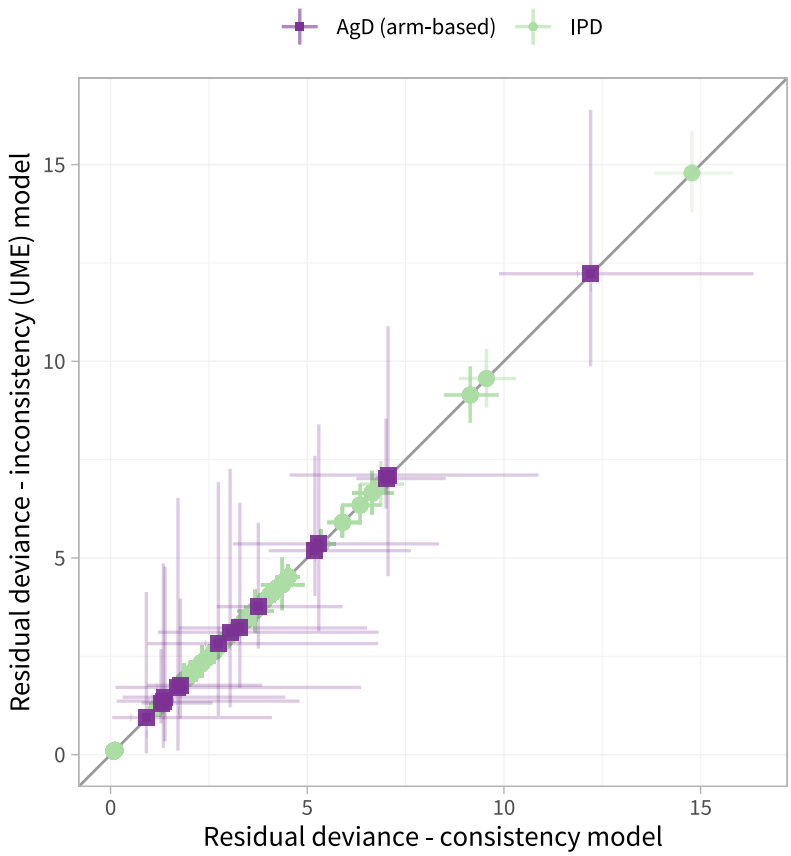

Figure A.2: Residual deviance contributions (posterior mean and 95% Credible Interval) under a fixed effect NMA model assuming consistency, and under a fixed effect NMA inconsistency (UME) model. Here, all of the posterior mean residual deviance contributions lie on the line of equality; all points fit equally well under either model.

A.5 Assessing the shared effect modifier assumption

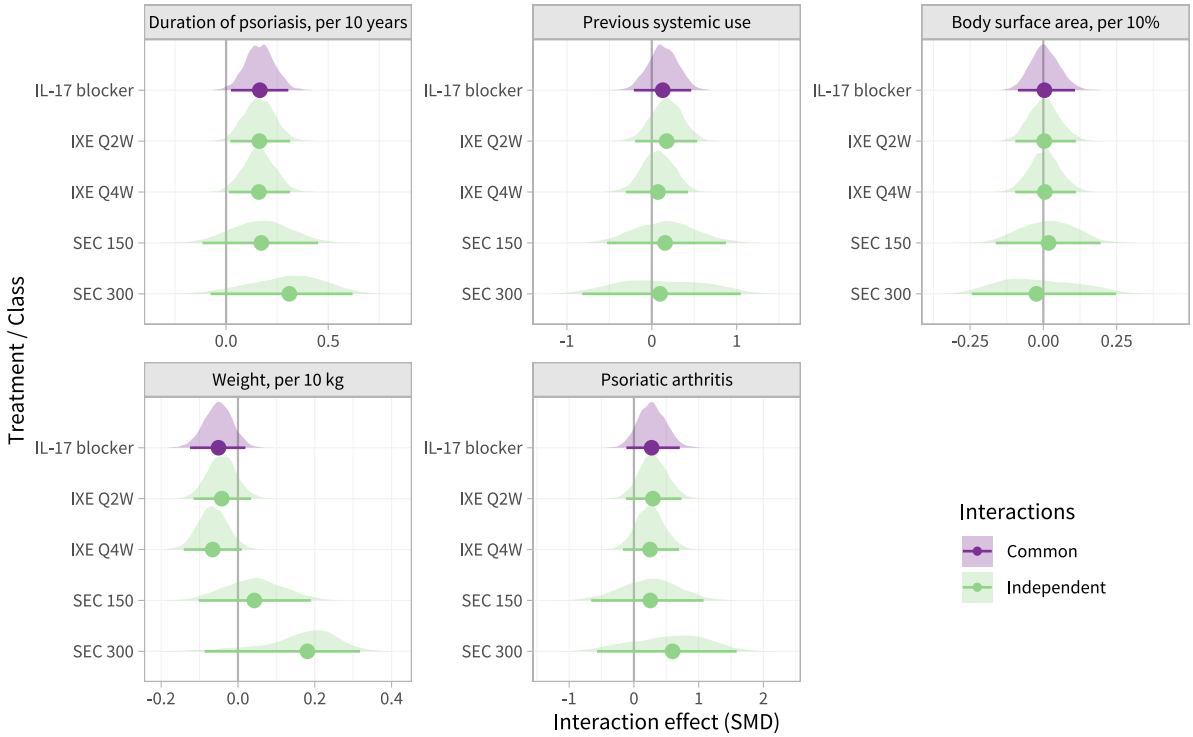

Figure A.3: Posterior estimates (median, 95% CrI, and density) of effect modifier interaction parameters from the ML-NMR model making the shared effect modifier assumption within the IL-17 blocker class for all covariates, and the ML-NMR models relaxing this to independent interactions for each covariate in turn.

Table A.10: Model fit statistics for the ML-NMR model making the shared effect modifier assumption within the IL-17 blocker class for all covariates, and the ML-NMR models relaxing this to independent interactions for each covariate in turn. p_D is a measure of the effective number of parameters. Residual deviance on 12,384 data points.

		Independent interactions for covariate				
	Common interactions	Duration of psoriasis	Previous systemics	Body surface area	Weight	Psoriatic arthritis
Residual deviance	8778.3	8780.9	8779.9	8780.9	8775.3	8780.6
p_D	36.6	38.9	38.6	38.7	39.6	38.5
DIC	8814.9	8819.9	8818.5	8819.6	8815.0	8819.1

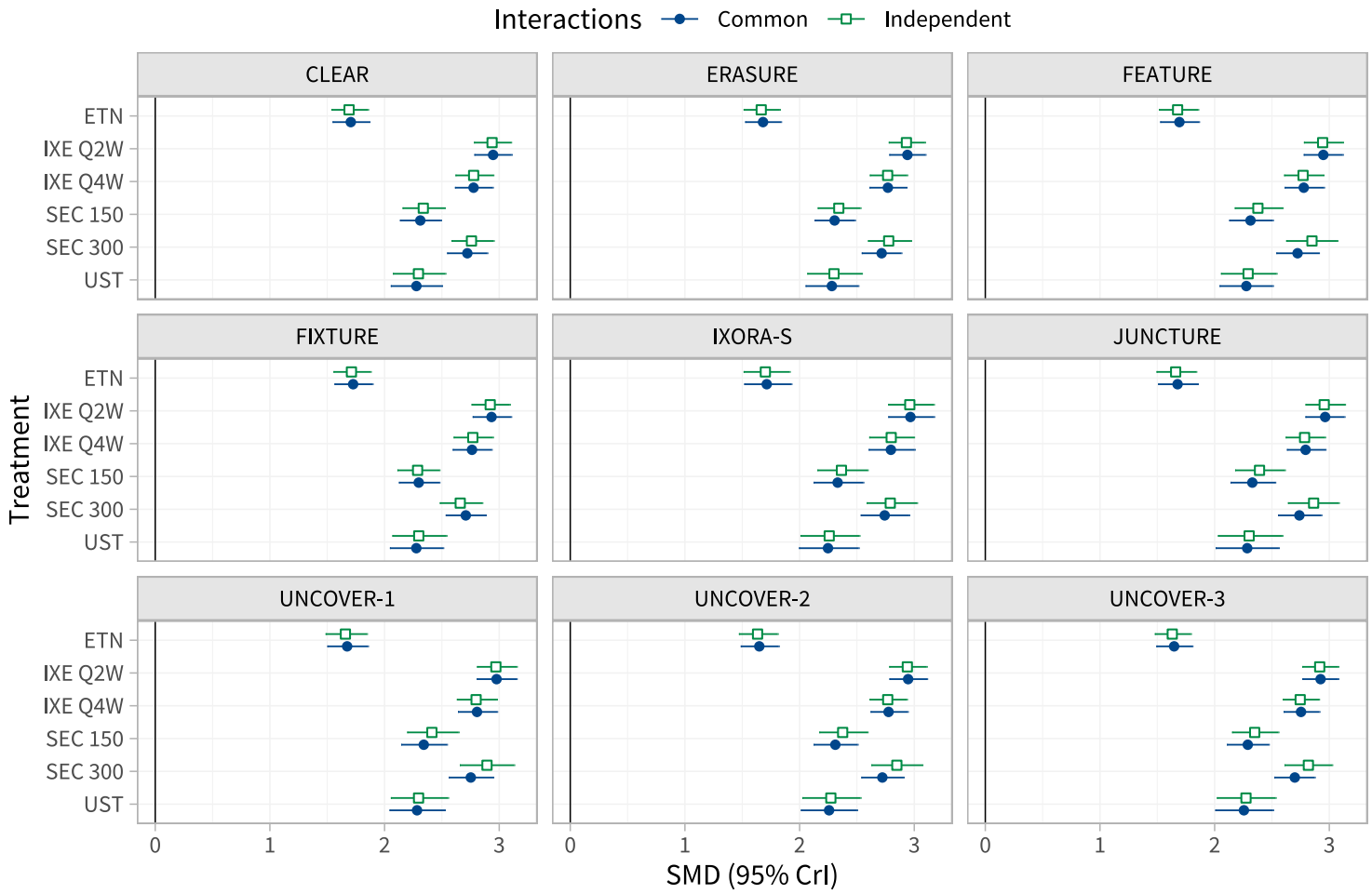

Figure A.4: Estimated population-average treatment effects (standardised mean differences) for each treatment vs. placebo in each study population, from the ML-NMR model making the shared effect modifier assumption within the IL-17 blocker class for all covariates, and the ML-NMR model relaxing this to independent interactions for weight.

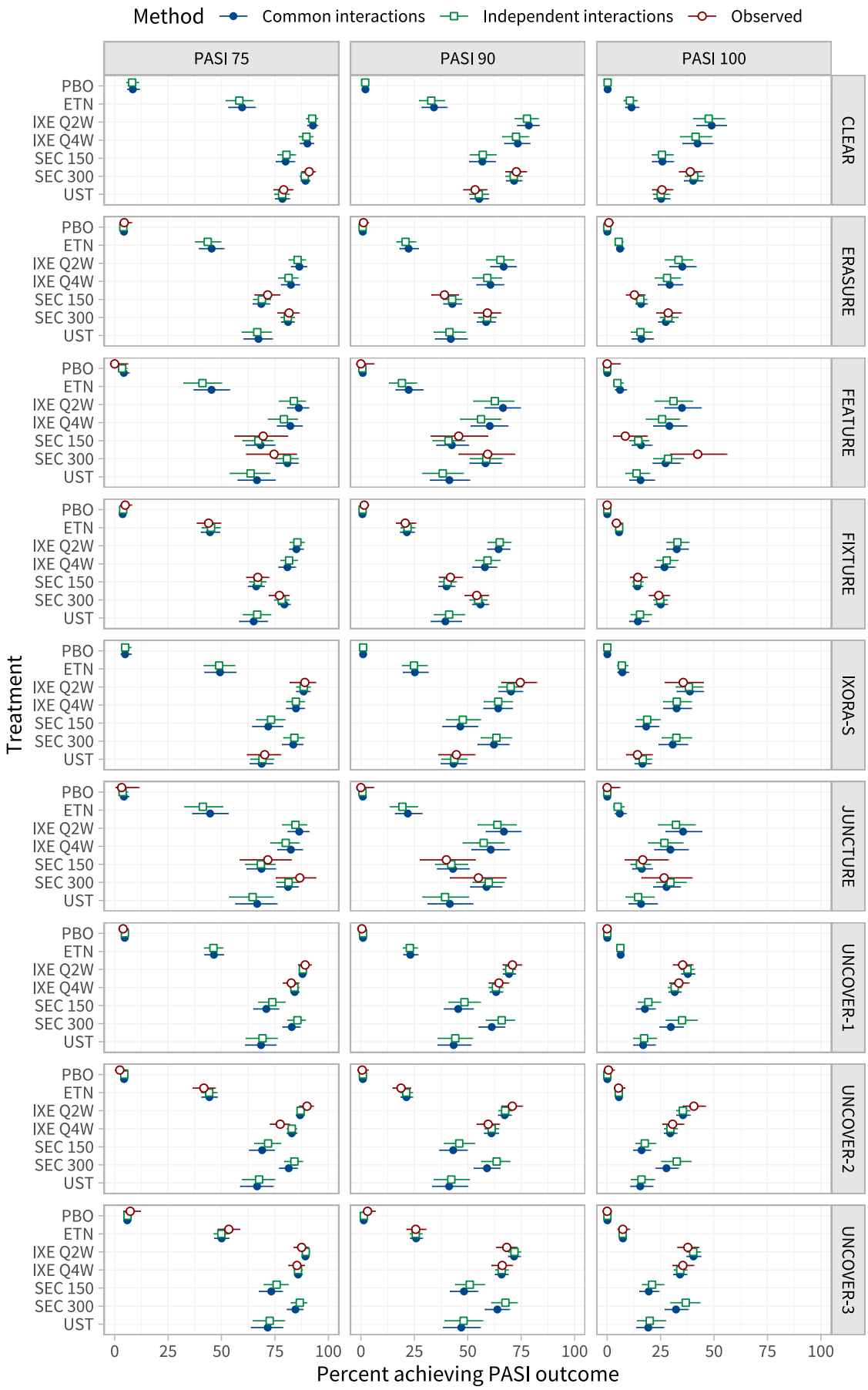

Figure A.5: Estimated percentage of individuals achieving each PASI endpoint on each treatment, in each study population, from the ML-NMR model making the shared effect modifier assumption within the IL-17 blocker class for all covariates, and the ML-NMR model relaxing this to independent interactions for weight.

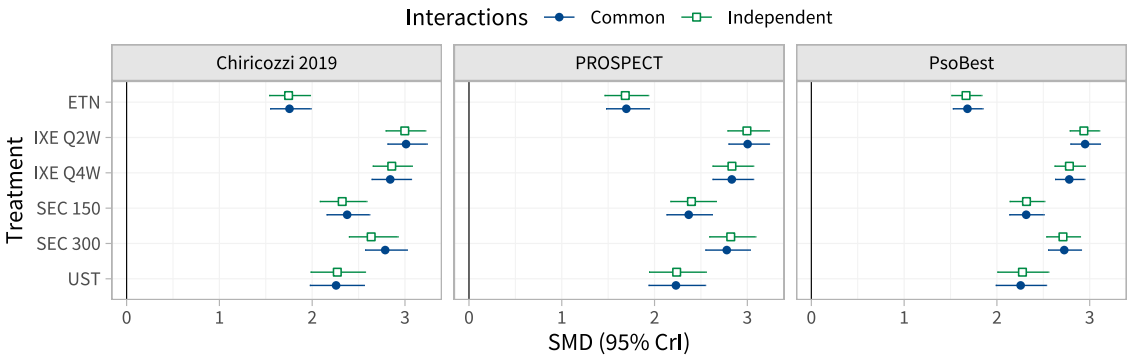

Figure A.6: Estimated population-average treatment effects (standardised mean differences) for each treatment vs. placebo in each external target population, from the ML-NMR model making the shared effect modifier assumption within the IL-17 blocker class for all covariates, and the ML-NMR model relaxing this to independent interactions for weight.

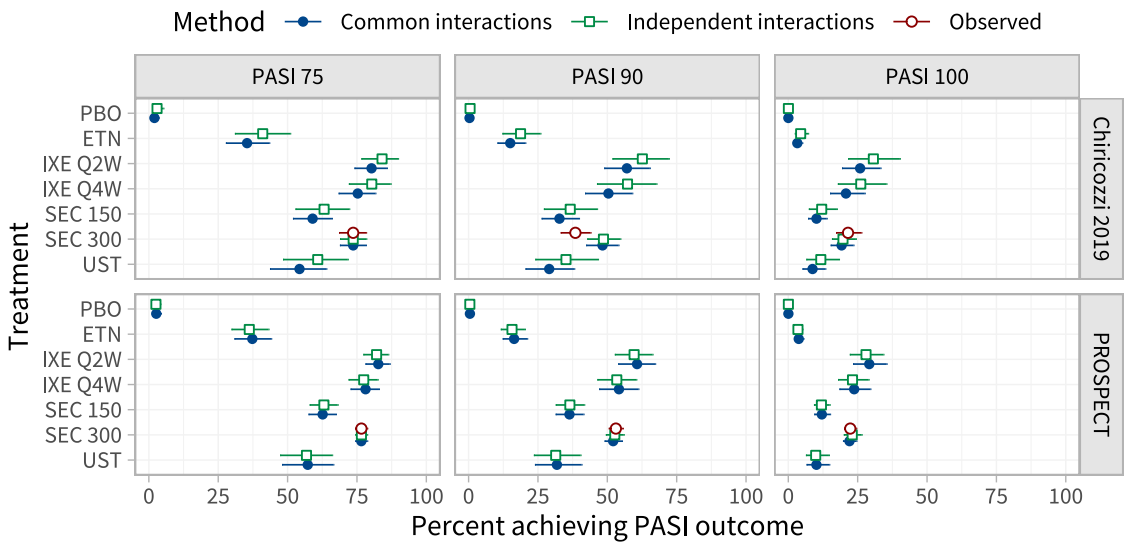

Figure A.7: Estimated percentage of individuals achieving each PASI endpoint on each treatment, in each external target population with information on response rates, from the ML-NMR model making the shared effect modifier assumption within the IL-17 blocker class for all covariates, and the ML-NMR model relaxing this to independent interactions for weight.