



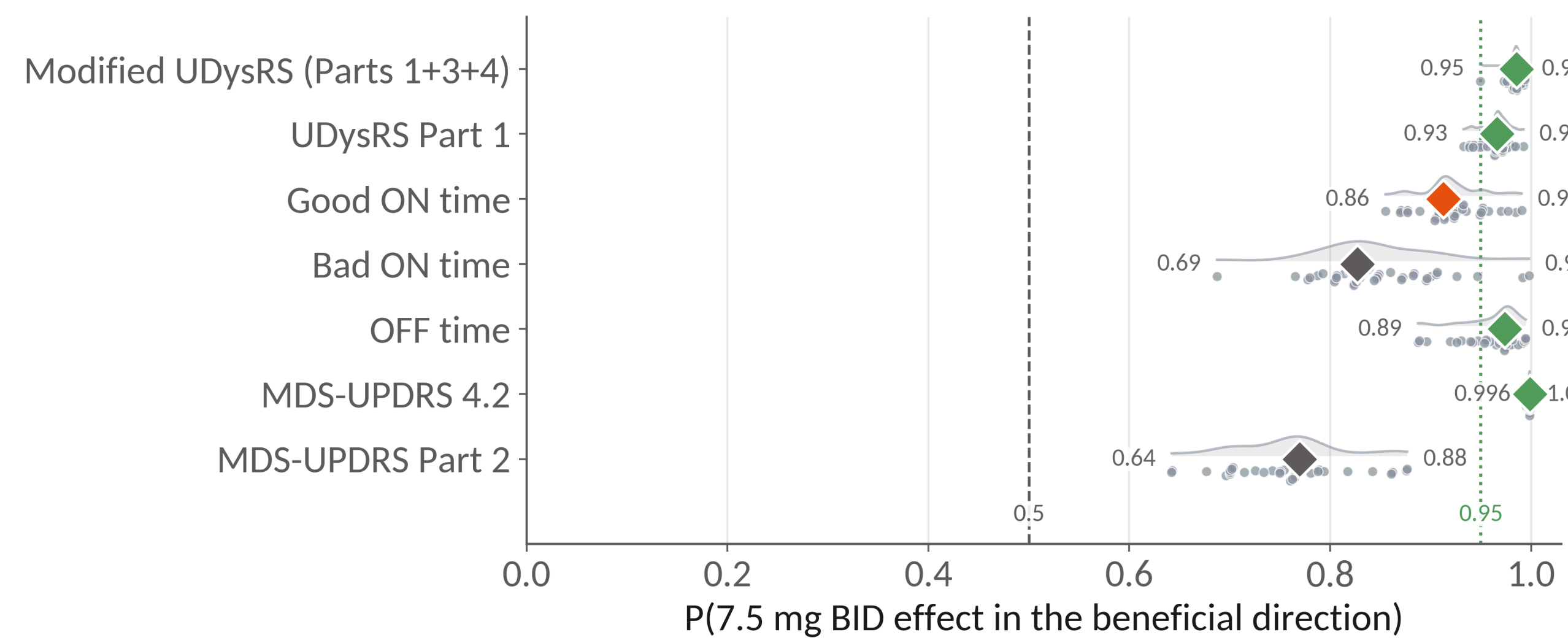
IRLAB

Investigating the robustness of a Bayesian meta-analysis of mesdopetam's anti-dyskinetic efficacy to different modelling choices

WERNER E., HANSSON F., WATERS S.

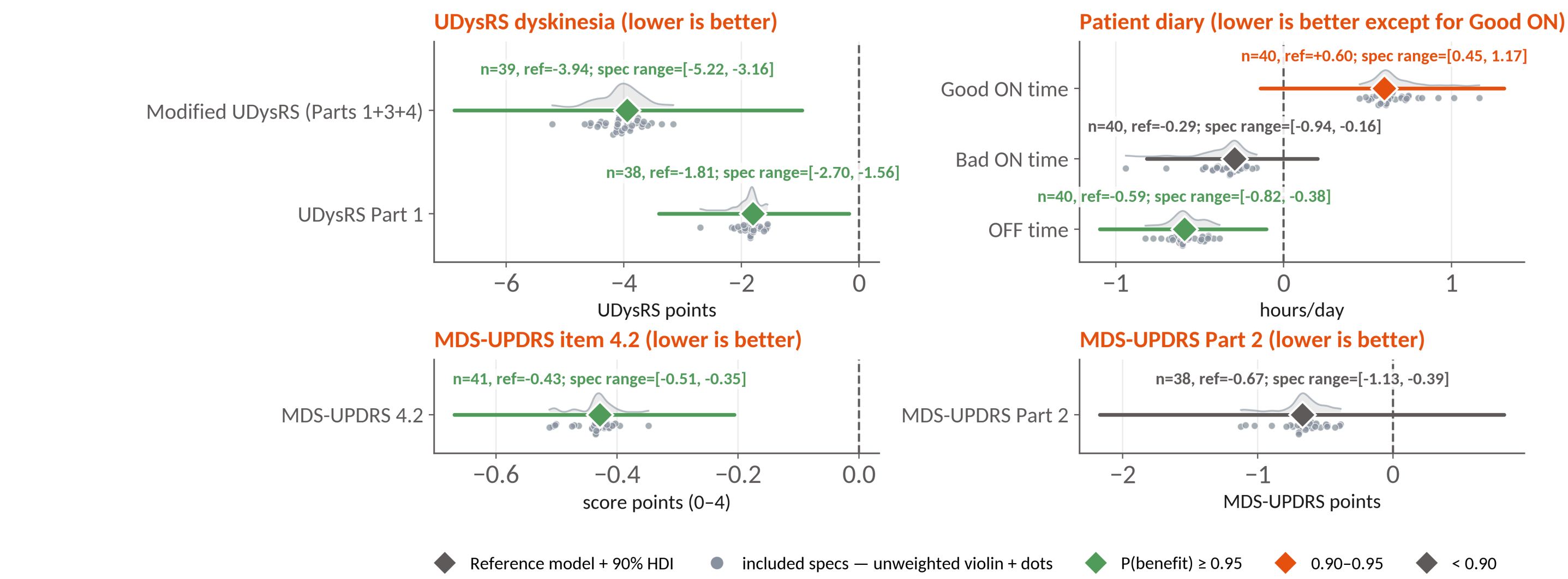
Integrative Research Laboratories (IRLAB), Gothenburg, Sweden · StanCon 2026

PROBABILITY OF BENEFIT



Posterior P(effect in the beneficial direction). Diamond = reference model. The raincloud (half-violin + one dot per specification) shows the full **unweighted distribution** of P across the included specifications. End labels give min/max P ; dashed/dotted lines mark $P = 0.50$ and $P = 0.95$.

EFFECT SIZE BY ENDPOINT



Reference-model effect with 90% HDI by measurement scale. Each effect is the **7.5 mg BID - placebo** difference in the endpoint's own raw units. The raincloud (half-violin + one dot per specification) shows the full **unweighted distribution** of specification means; labels give the number of specifications, the reference **mean effect size**, and the range over included specifications.

INTRODUCTION

Mesdopetam (IRL790) is a dopamine D_3 receptor antagonist in clinical development for **levodopa-induced dyskinesia** in Parkinson's disease (**PD-LID**). PD-LID, which can be described as involuntary movements triggered by anti-parkinsonian medication, is a disabling condition that worsens as the disease advances, and there is a great need for improved medical therapies. The fluctuating nature of PD-LID and the partly subjective nature of the associated functional impairment make the evaluation of therapeutic interventions challenging, although several measurement instruments are available. The clinical efficacy of mesdopetam has been evaluated in a Phase IIa study (NCT03368170) and a Phase IIb study (NCT04435431)^[1], both randomized controlled trials comparing mesdopetam against placebo. While both trials indicated clear anti-dyskinetic effects based on secondary endpoints, the primary endpoints, UDysRS for Phase IIa and Good ON time for Phase IIb, were not met. A Bayesian meta-analysis of the results was previously presented as part of a poster at the 2024 MDS Congress^[2]. That work included a single Bayesian meta-analysis; here we investigate how results vary under a large number of plausible model specifications.

A joint analysis of the two studies must address the fact that they differ in many important respects, including the patient population, the duration of the study and the dosing regimen. Further, even for a single study there are a large number of researcher degrees of freedom in terms of e.g. the statistical model used for each endpoint, the selection of a prior distribution, and the handling of missing data and dropouts. In this post-hoc analysis, we construct and sample from a large number of Bayesian models to investigate how sensitive our efficacy estimates are to different modelling choices. Models are implemented in PyMC and sampled using Nutpie.

MODEL UNIVERSE

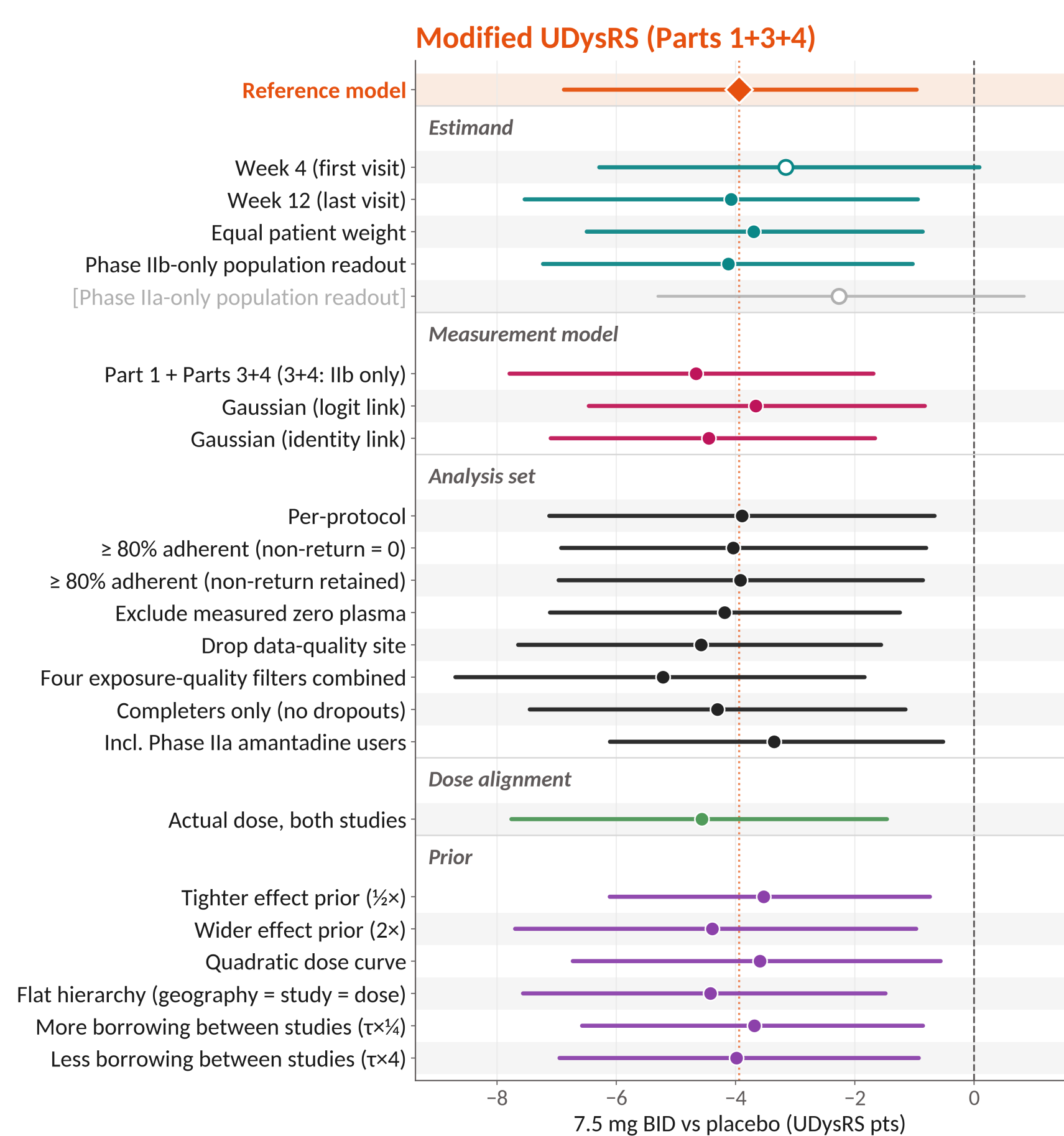
Each endpoint is modelled as a hierarchical regression on **latent severity**. Each visit- t severity carries the latent baseline b_i of patient i forward and adds a **change** predictor η_{it} . The latent scale is mapped to the observed endpoint y_{it} through the measurement model D .

$$\begin{aligned} y_{i0} &\sim D(b_i), & y_{it} &\sim D(b_i + \eta_{it}) \\ b_i &\sim \mathcal{N}(\mu_b, \sigma_b) \\ \eta_{it} &= \alpha + \beta b_i + \delta_{d(i)l(i)} + \gamma_t + (\delta\gamma)_{d(i)t} + c_{k(i)} + s_{l(i)} + u_i \\ \delta_{dl} &\sim \mathcal{N}(\bar{\delta}_d, \tau_{\text{dose}}), & \delta_{0l} &= 0, & \tau_{\text{dose}} &\sim \mathcal{N}^+(0, \sigma_\tau) \end{aligned}$$

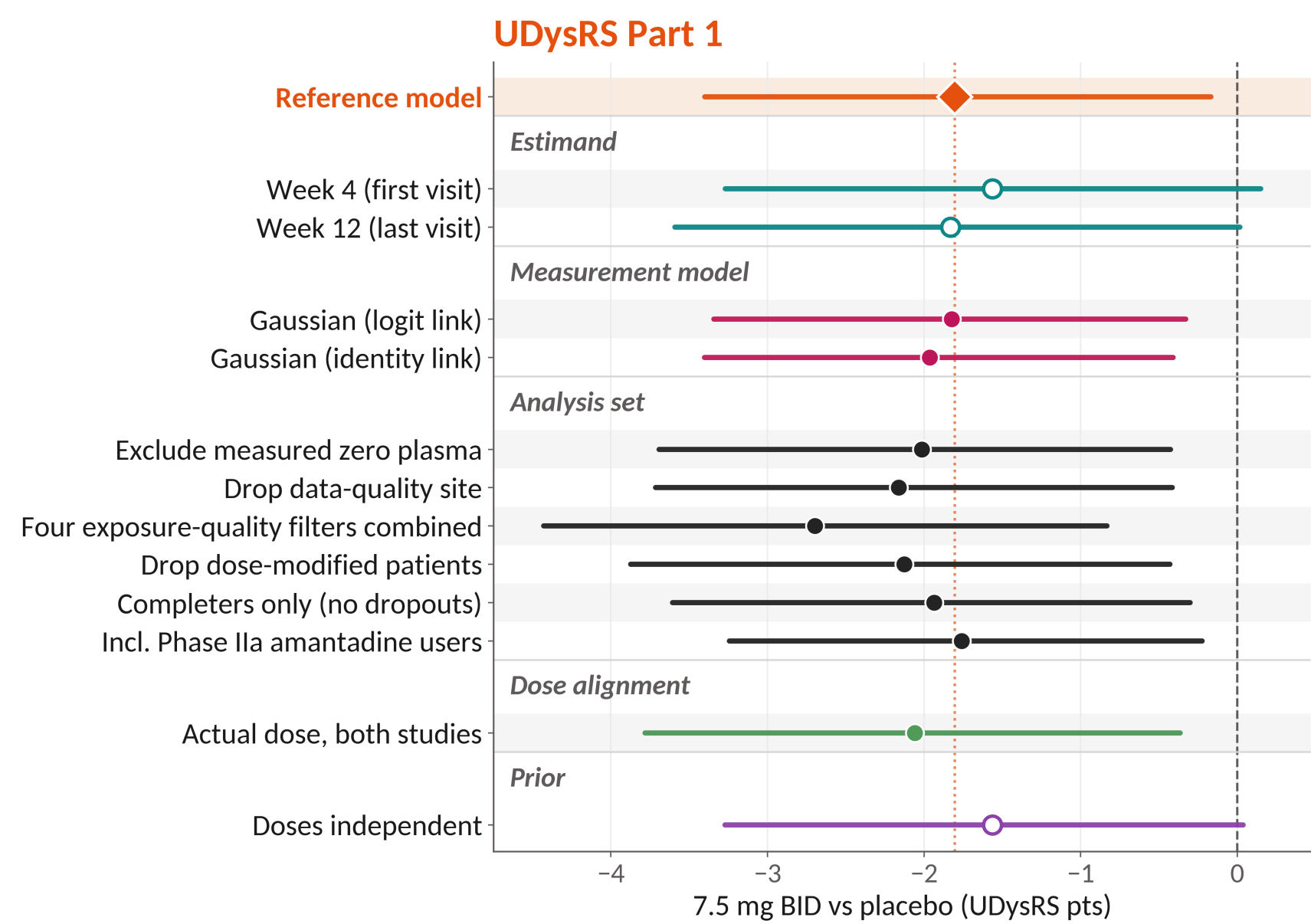
The latent change predictor η_{it} has terms for latent baseline severity (βb_i), a study-specific dose effect ($\delta_{d(i)l(i)}$), visit (γ_t) and dose $\times$ visit terms, country/study offsets and a patient intercept; $d(i)$, $l(i)$, $k(i)$ map patient i to dose arm, study and country. δ_{dl} is the default hierarchical study $\times$ dose interaction: a shared dose-response curve $\bar{\delta}$ plus study deviations, with learned SD τ_{dose} ; its prior scale σ_τ controls borrowing ($\tau_{\text{dose}} \rightarrow 0$ gives full pooling).

For each endpoint we define a **reference model**, the diamond in the above plots. Bounded item sums use a beta-binomial distribution as the measurement model, the 0-4 UPDRS item uses ordinal logit, and diary hours use a joint generalized Dirichlet-multinomial composition. The reference model assumes random-walk dose and visit curves and regularizing priors, and is evaluated as the **7.5 mg BID versus placebo** contrast time-averaged over Weeks 4, 8 and 12 on the pooled two-study population. We then refit a curated sensitivity grid, varying one main modelling choice at a time where possible. The **measurement family** D : Normally distributed with a logit or an identity link; beta-binomial models for bounded item sums; binomial, ordinal logit **and** probit for the single 0-4 item; and Dirichlet, Dirichlet-multinomial, logistic-normal, generalized Dirichlet and generalized Dirichlet-multinomial families for the diary endpoints. We compare a number of different priors for the **dose-response relationship** $\bar{\delta}$: random walk with diminishing prior step SDs (the reference), second-order random walk, Gaussian process, monotone, Emax, quadratic or fully independent, with scales calibrated to the same average prior SD across active doses. The **visit-effect curve** γ is likewise a random walk or independent. We also vary **prior scales**: halving or doubling the overall effect scale, widening dose $\times$ visit interactions, tightening or loosening country, study, patient and borrowing scales, and removing the extra shrinkage on country, study, and dose $\times$ visit effects. **Dose alignment** d is taken as-randomized or swapped for the dose actually received in Phase IIa, Phase IIb or both. **Analysis-set** filters restrict the sample by per-protocol status, dose-modification, site quality, compliance, measured-zero plasma exposure and their four-filter combination, with a complementary add-back that restores the Phase IIa amantadine users. For **missing data** we compare available-case against a completers-only sample. Finally the **estimand** varies the *readout* of the contrast against placebo: pooled or Phase IIb-only population (Phase IIa-only population is shown in the plots below, but excluded from the ensemble), patient-visit or equal-patient weighting, and time-averaged over Weeks 4, 8 and 12 or read at Week 4 or Week 12. Regularizing priors are specified on each model's latent scale, by default with stronger shrinkage on country, study, and dose $\times$ visit effects than on intercept, dose, visit, and patient effects.

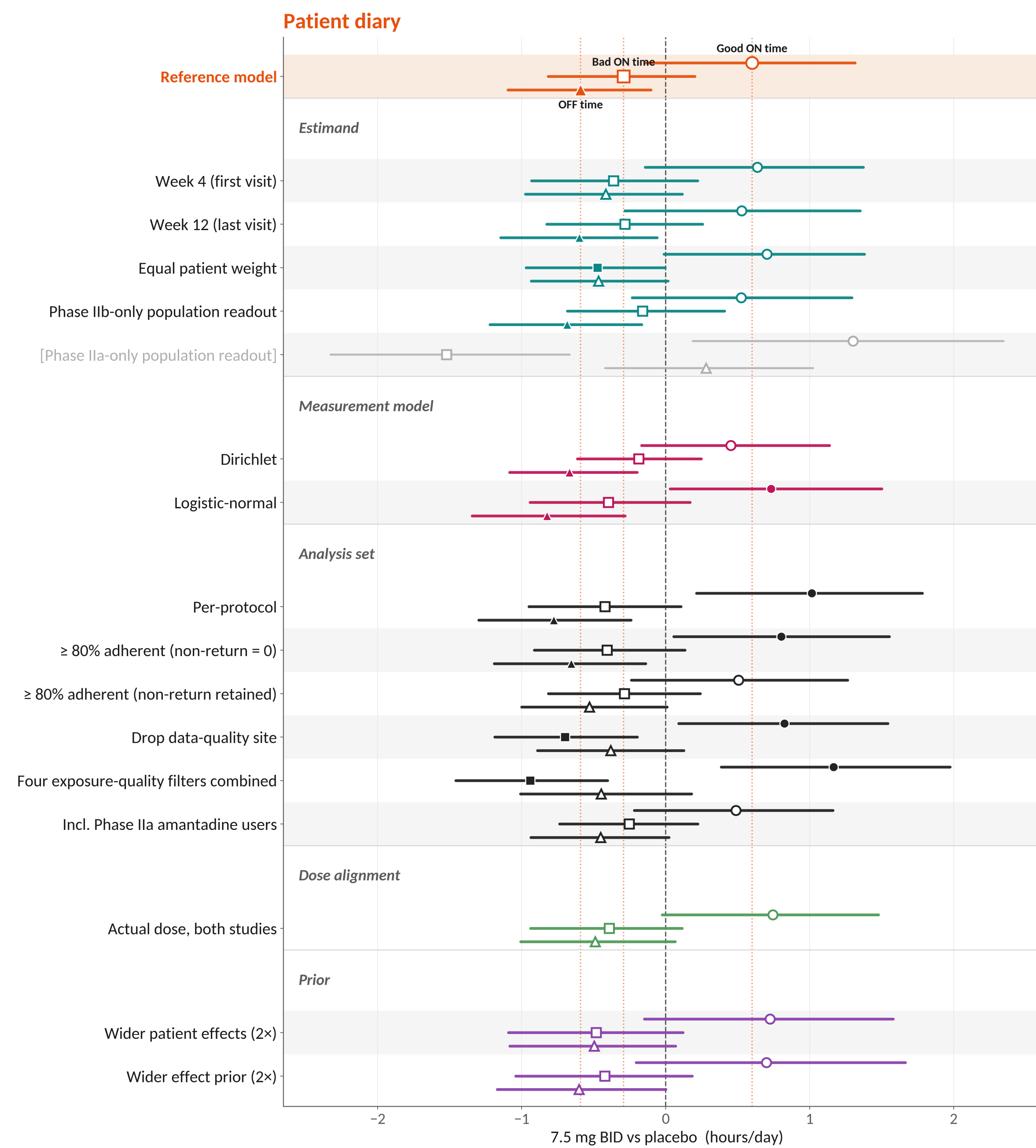
PER-ENDPOINT FORESTS — SELECTED SENSITIVITY ANALYSES



Modified UDysRS, Parts 1+3+4 summed (0-88, baseline average 37.6) — 22 items scored 0-4. Part 1 is historical dyskinesia complications; Parts 3+4 are examiner-rated ON-state dyskinesia. The default is beta-binomial (N=88); a measurement variant drops unreliable Phase IIa Parts 3+4. Results are clearly beneficial for almost all included models. Actual-dose coding, combined exposure-quality filtering and a wider effect prior strengthen the effect; including IIa amantadine users or a tighter prior attenuate it.



UDysRS Part 1 — historical disability (0-44, baseline average 19.4) — 11 historical dyskinesia complication items. Default: beta-binomial (N=44). The effect is clearly beneficial for most specifications. Analysis-set and actual-dose variants move the estimate the most, with measurement-model, estimand and prior rows contributing more modestly.



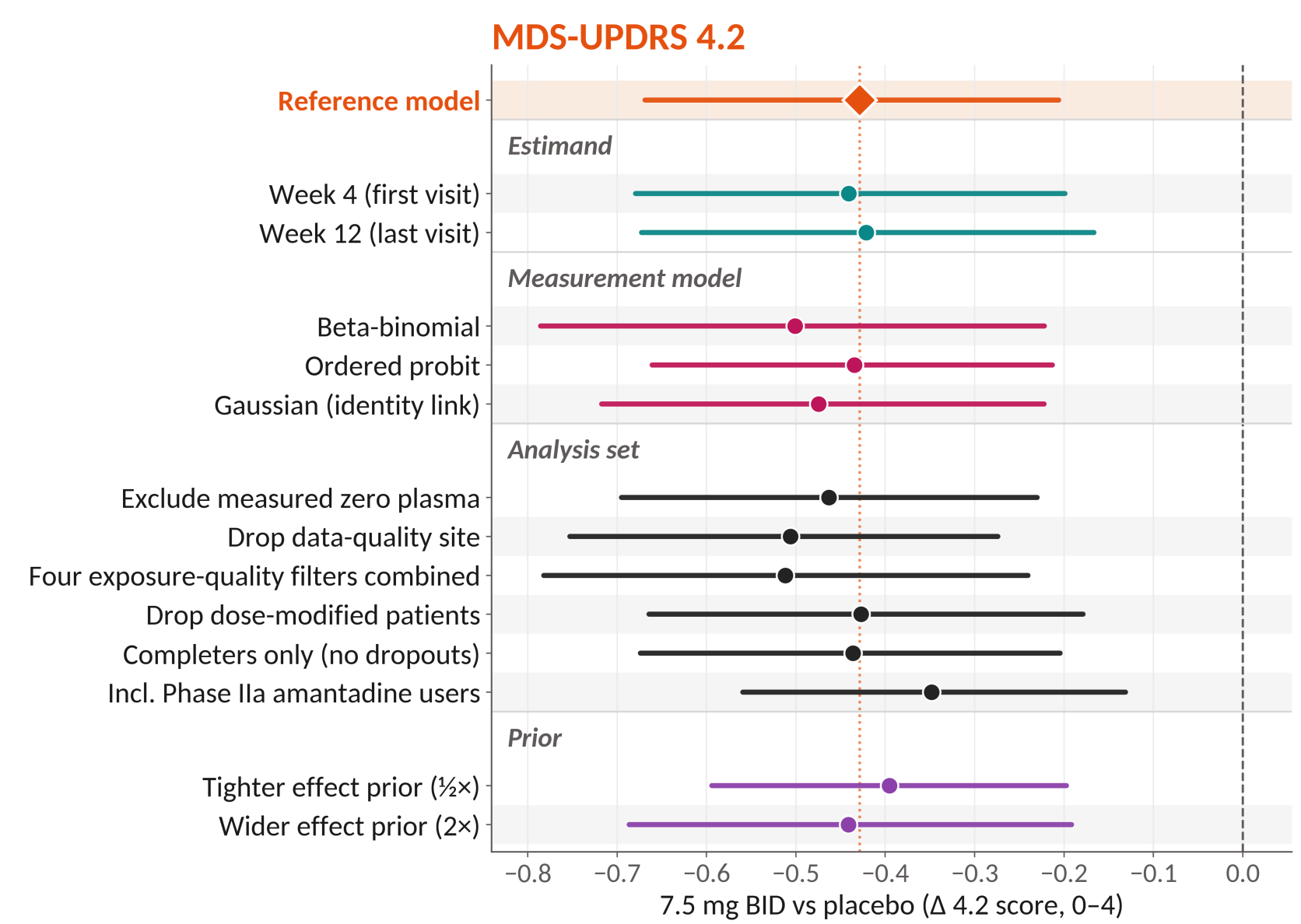
Patient diary — Good ON time (baseline average 6.7), Bad ON time (5.6) and OFF time (3.9) hours per day, fit jointly. Patients log each half-hour of the day as Good ON time, Bad ON time, OFF time, or asleep; the four sum to 24 h, forming a daily composition (we report the three waking states). The default measurement model is a generalized Dirichlet-multinomial on that composition, with sleep as the reference state. The results are quite sensitive to both the measurement model and the analysis set (stricter compliance checks tend to have stronger effects). Prior variants are less influential than the measurement and analysis-set choices; the joint model and repeated diary days yield a more informative likelihood than the other endpoints.

CONCLUSIONS

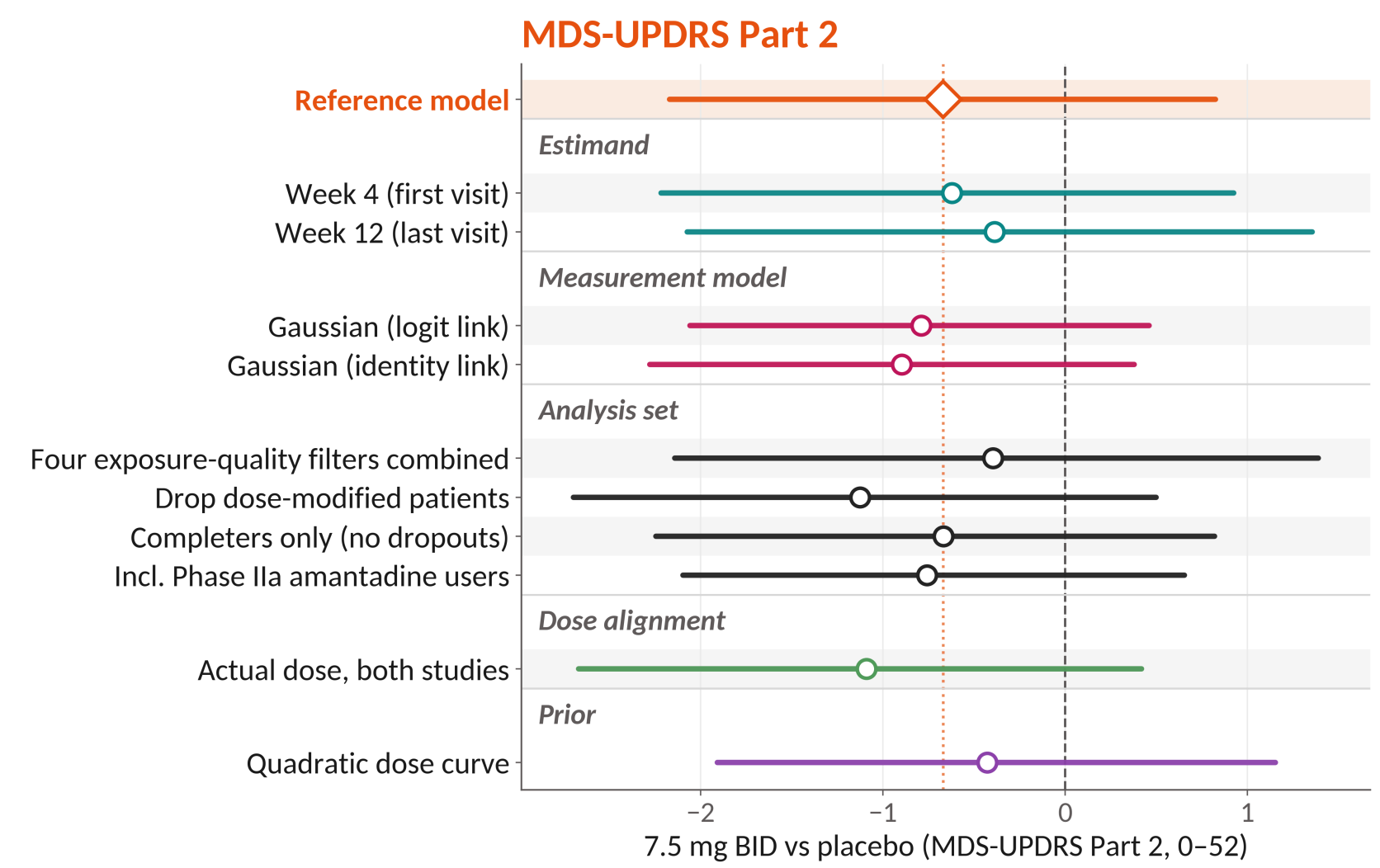
For the range of instruments used to assess dyskinesia in the two trials, estimates generally favour an anti-dyskinetic effect of mesdopetam across the included model specifications. The comprehensive, standard dyskinesia assessment scale, UDysRS, is clearly improved versus placebo under almost all included specifications, both in terms of the total sum of parts 1, 3, and 4 assessing dyskinesia severity and disability, and part 1 assessing disability over the past week. Likewise, UPDRS 4.2, a single item capturing functional disability, is consistently improved versus placebo, with only a modest dependence of the effect size on the measurement model. Motor diary endpoints are very sensitive to both the measurement model and data-quality filtering, with OFF time showing a benefit under most included specifications, and Good ON time under strict data filtering assumptions. This may be related to the well-recognized shortcomings in the assessment of motor states

by home diaries, where in particular assessment of Bad versus Good ON is challenging, while OFF can be more accurately assessed^[1]. Importantly, MDS-UPDRS Part 2 (motor experiences of daily living), included as a safety check, shows no sign of worsening under any included specification.

Taken together, these analyses, combining data from two Phase II trials, each applying somewhat different designs and with some technical challenges mainly related to the inherent difficulty of reliably quantifying levodopa-induced dyskinesias, indicate consistent anti-dyskinetic effects in terms of both objective and subjective assessments. Furthermore, these effects are not accompanied by detected worsening in motor experiences of daily living. Hence, mesdopetam could be a meaningful therapeutic option as add-on to levodopa for patients with advanced Parkinson's disease and disabling dyskinesia.



MDS-UPDRS 4.2 — functional impact of dyskinesia, a single 0-4 item (baseline average 2.2). A clinician's rating of how much dyskinesia interferes with daily function. The default measurement model is an ordinal logit, since a single under-dispersed item is misfit by a beta-binomial. The effect size is fairly stable across likelihoods and the sign never changes across the included specifications.



MDS-UPDRS Part 2 — motor experiences of daily living, 13 items (0-52, baseline average 16.3). This endpoint is a **safety** rather than a dyskinesia measure: How does mesdopetam affect the motor experiences of daily living? The default measurement model is a beta-binomial with N=52. Under all specifications, the mean effect is a tiny improvement and the 90% HDI straddles zero — **no sign of motor worsening**.