



Making the Most of Indirect Evidence: IPD Network Meta-Analysis for Preclinical Networks

Ezgi Tanriver-Ayder, Adetayo Kasim
Oncology Discovery Statistics, GSK
NCS 2026

How do we compare treatments never tested together?

Question:

Drug A and Drug B were never tested in the same study. Both were compared against a common group in separate studies. How do we reliably estimate the effect of A vs B?

1

Ignore it

Skip the indirect comparison entirely. Common in nonclinical practice but wastes available evidence.

2

Ad-hoc adjustment

Subtract or scale estimates informally. No framework for multiple comparisons or heterogeneity.

3

Network meta-analysis (NMA)

Formally synthesise direct and indirect evidence. Individual data NMA extends this to the individual donor level.

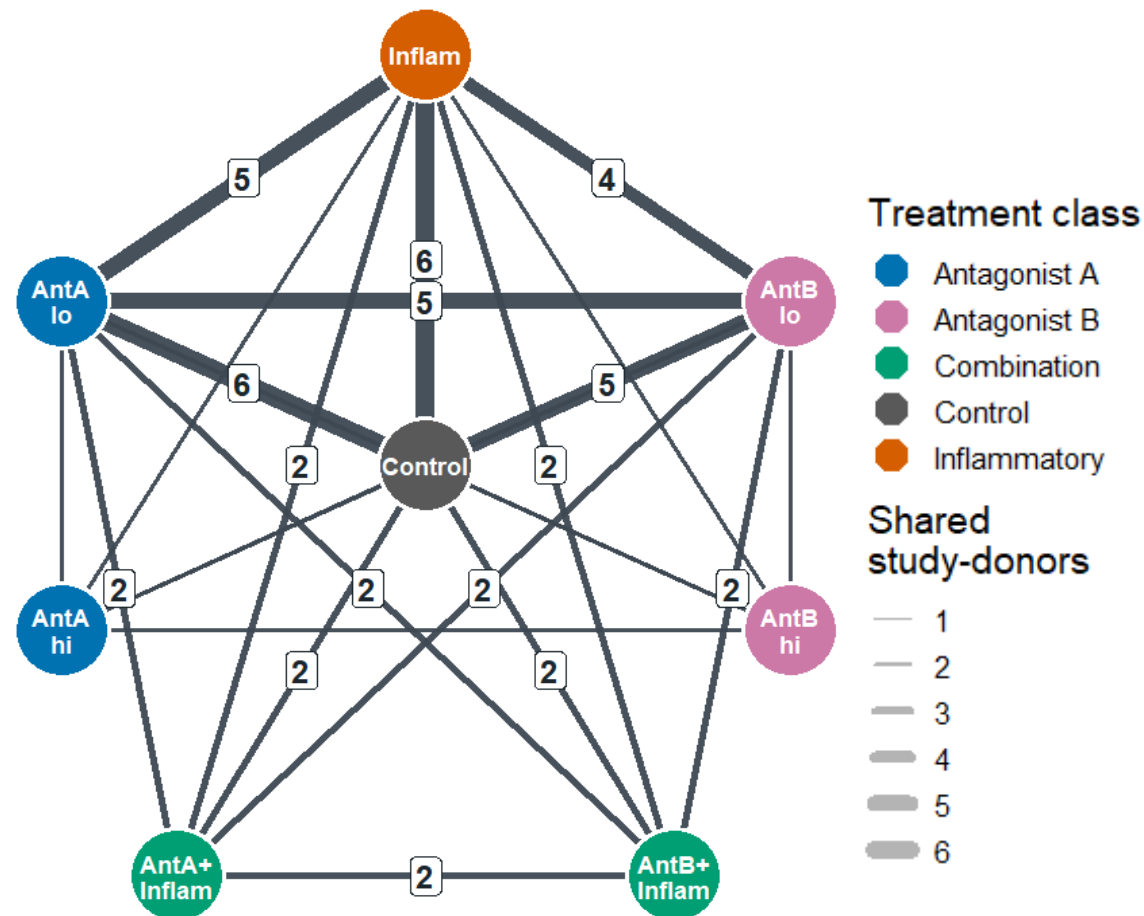
Challenges specific to preclinical data

- Sparse, unbalanced networks:
 - Not all treatments tested in each experiment.
 - Incomplete treatment overlap means most comparisons rely on indirect evidence
- Non standardised experimental protocols:
 - Differences in lab protocol and experimental conditions contribute to within experiment variability that varies across studies
- Small sample sizes:
 - Nonclinical networks are often based on limited number of replicates, experiments and donors, making between donor variability difficult to estimate reliably, especially in sparse networks
- Donor level biological variability:
 - The same treatment can produce substantially different responses across tissue donors - a key source of between-donor heterogeneity

Motivating example: Ex-vivo human data

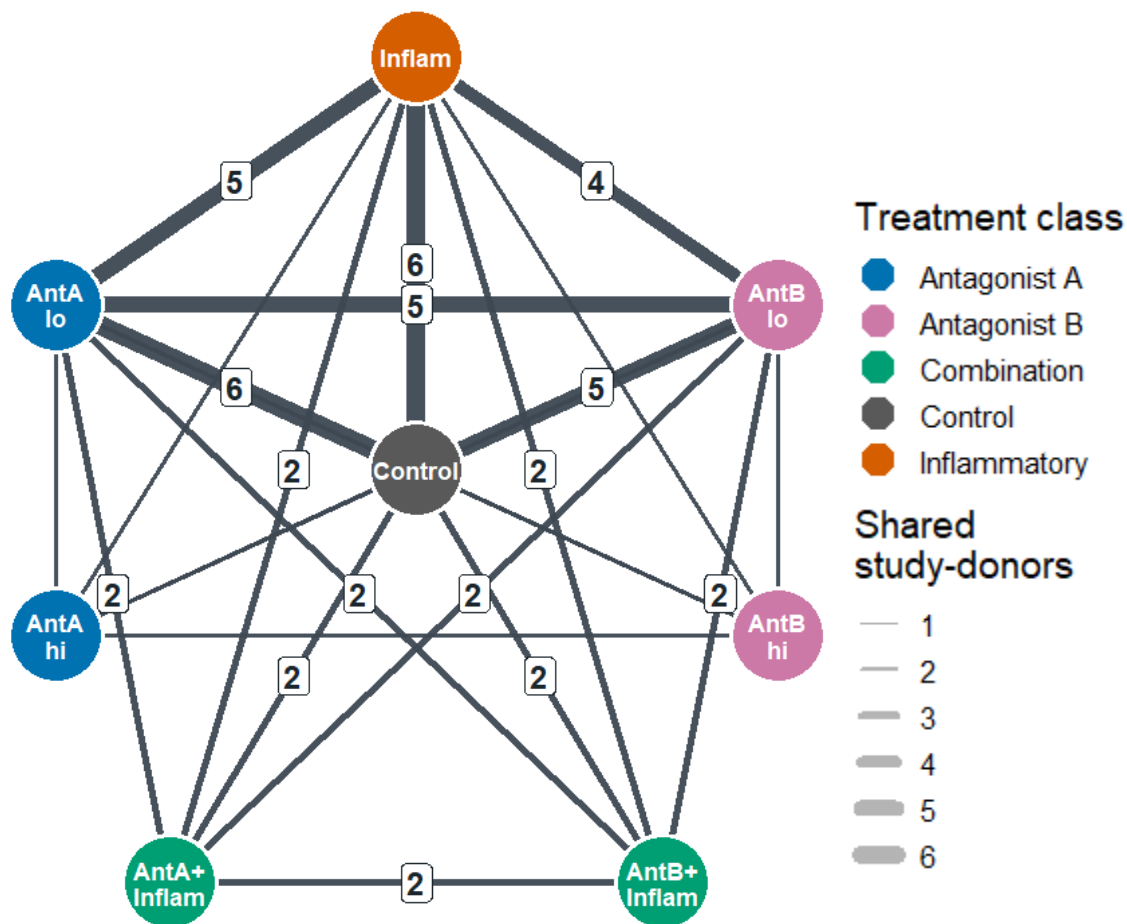
Study Design:

- 7 ex-vivo human study donors across 5 parent studies
- Response: Calcium imaging (P2/P1 ratio)
- 8 treatment groups: control, 2 antagonist (low, high), inflammatory sensitiser, 2 combination arms
- 6 replicate cells per treatment per study donor
- Note that not all treatments are tested in every study donor, mostly indirect evidence -> sparse network



*Edge width = number of study donors with direct comparison

Motivating example: Ex-vivo human data



*Edge width = number of study donors with direct comparison

- High dose and combination arms are never in the same study donor. No direct evidence.
- Indirect comparisons are estimated through shared comparators — Control (present in all donors) and the low dose arms, which bridge across studies.
- Only individual data NMA borrows through all network paths while preserving within-donor structure; giving the most precise estimates.

Three methods: Same data, different use of donor structure

	Individual Data NMA (IPD-NMA)	Aggregate Data NMA (AD-NMA)	Bucher Indirect Treatment Comparison (Bucher-ITC)
Unit	Each study-donor (7 units)	Each parent study (5 units)	Study-donor (two-step formula)
Evidence	Indirect/direct	Indirect/direct	Indirect/direct
Donor pairing	Preserved	Pooled first	Partial
Model	netmeta on study-donor contrasts	netmeta on aggregated contrasts	$ITC(A \text{ vs } C) = D(A-B) - D(C-B)$
Key notes	<i>Within-donor contrasts cancel baseline differences. Full network borrowing.</i>	<i>Loses donor level signal.</i>	<i>Less efficient and unstable under high donor variability.</i>

What the simulation study asks?

1. *Does IPD-NMA produce more precise and better calibrated estimates than AD-NMA and Bucher ITC for indirect comparison?*

Why useful: If aggregation discards useful within donor information, we expect individual level NMA to show tighter and more reliable estimates.

2. *Under which network and design conditions is the individual level data advantage largest – and when does it disappear?*

Why useful: Testing this guides practical decision-making.

3. *How robust are the findings to donor heterogeneity, replicate count, and within study variability?*

Why useful: These three quantities vary across preclinical studies – important for assessing generalisability.

Simulation designs & data generating mechanism

- S1 – Actual study conditions: ($\tau=0.12$ | $\sigma=0.12$ | $n=6$)
- S2 – High study-donor variability: ($\tau=0.20$ | $\sigma=0.12$ | $n=6$)
- S3 – Low study-donor variability: ($\tau=0.05$ | $\sigma=0.12$ | $n=6$)
- S4 – Sparse data: Fewer treatment overlaps
- S5 – Dense data: All treatments in all donors
- S6 – Fewer replicates: ($\tau=0.12$ | $\sigma=0.12$ | $n=3$)
- S7 – Many replicates: ($\tau=0.12$ | $\sigma=0.12$ | $n=12$)
- S8 – High experimental noise: ($\tau=0.12$ | $\sigma=0.20$ | $n=6$)

*Each simulated dataset is analysed by all 3 methods (individual level NMA, aggregate level NMA and Bucher ITC) simultaneously.

$$y(s,i,k) = \mu_v + \delta_k + u_s + \varepsilon$$

μ_v = vehicle baseline

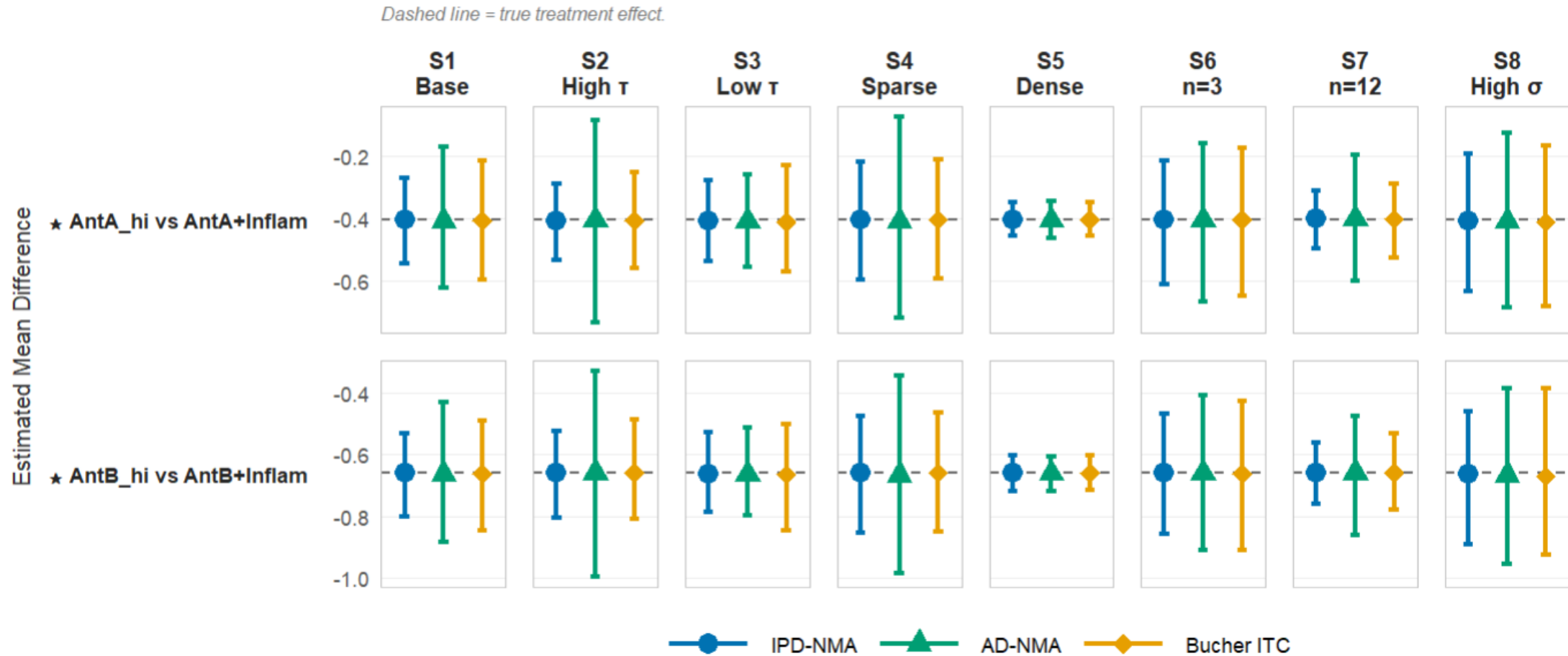
δ_k = treatment effect

$u_s \sim N(0, \tau^2)$ = study-donor random effect (between unit heterogeneity)

$\varepsilon \sim N(0, \sigma^2)$ = residual variability (experimental noise)

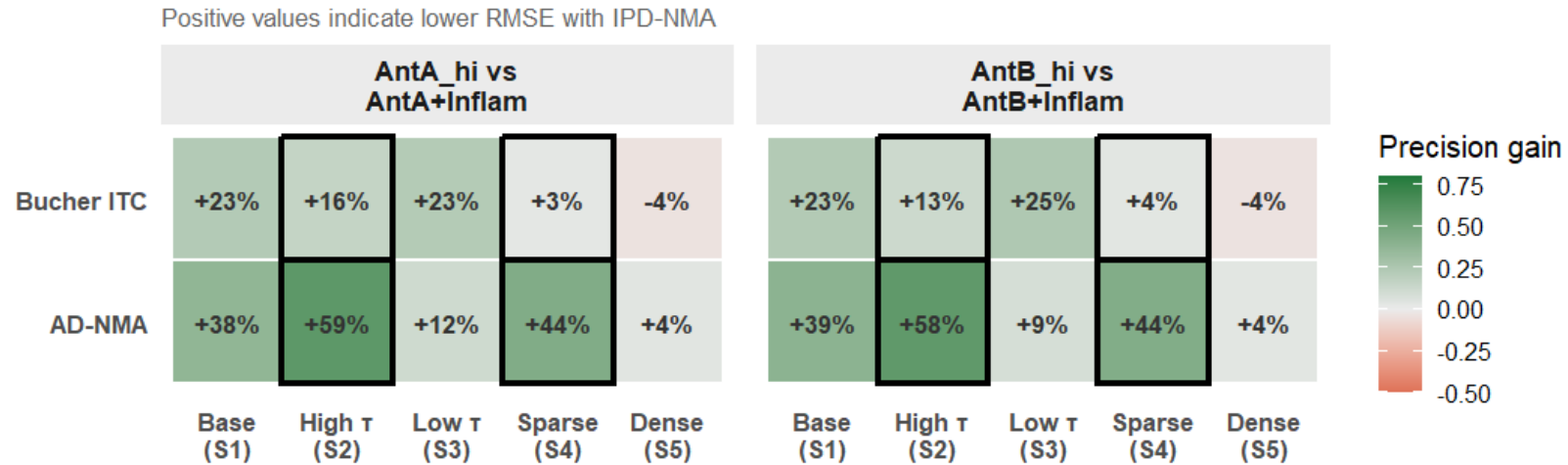
*All the above parameters are true effects estimated from empirical NMA on real study data

IPD-NMA produces most precise intervals for truly indirect comparisons



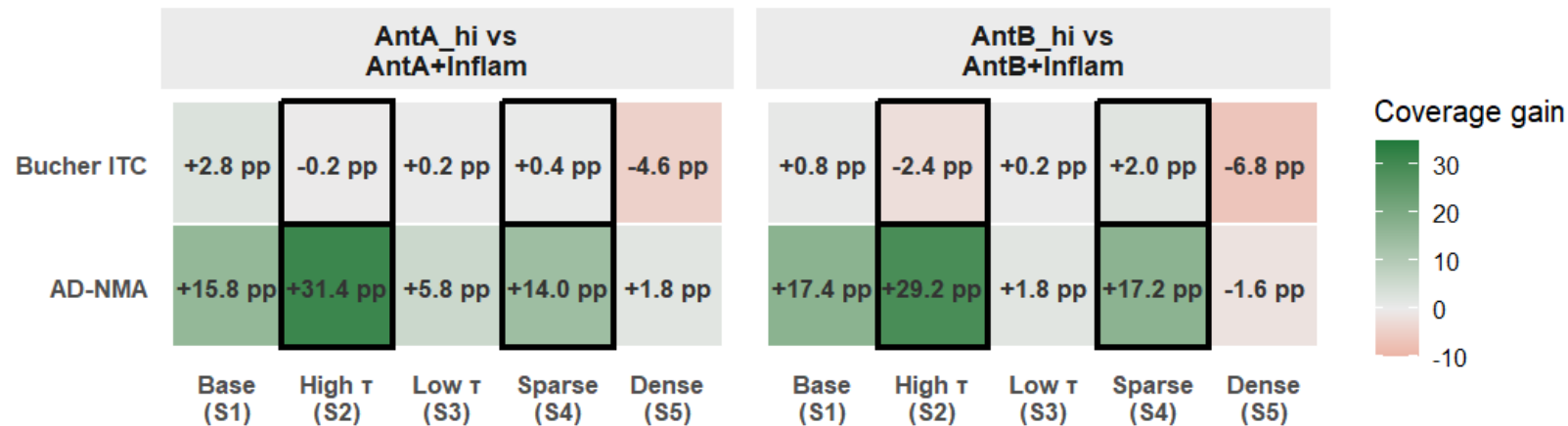
- IPD-NMA lands on the same estimate as other methods — but with much less uncertainty, especially in high heterogeneity, re replicates and sparse settings giving it more power to detect differences.

Under which network/design is the individual level data advantageous?



Coverage gain using IPD-NMA

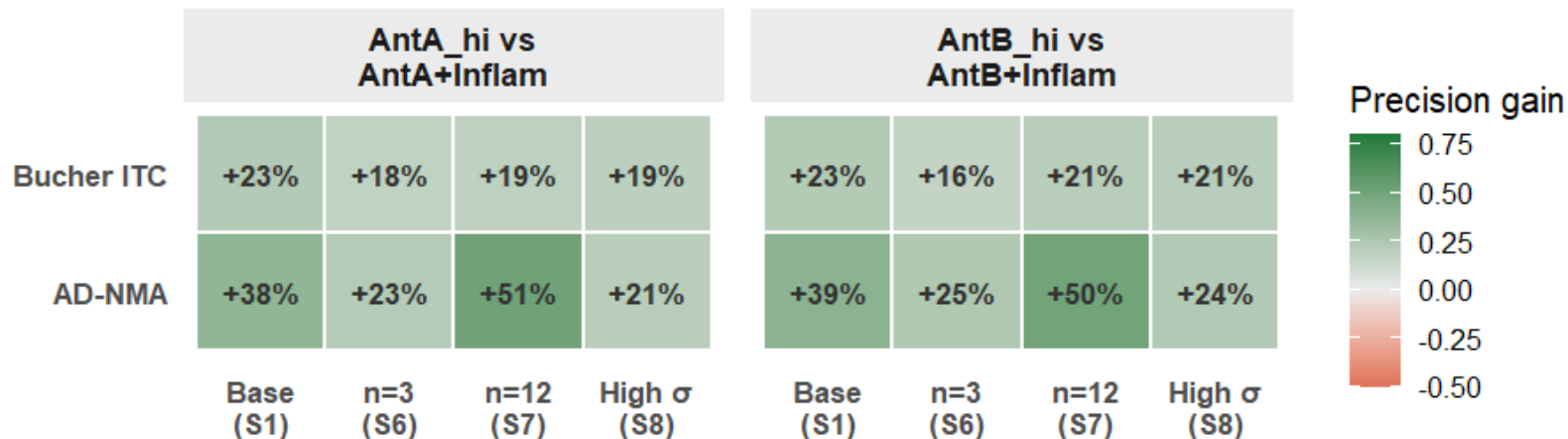
Positive values in percent point (pp) indicate more reliable coverage with IPD-NMA



- IPD-NMA is most valuable when:
 - donor effects are heterogeneous,
 - networks are sparse with indirect evidence.
- In these settings, AD-NMA was far less accurate, and its uncertainty ranges were unreliable
- Bucher ITC was closer to IPD-NMA than AD-NMA, but IPD-NMA still showed higher precision and reliability.

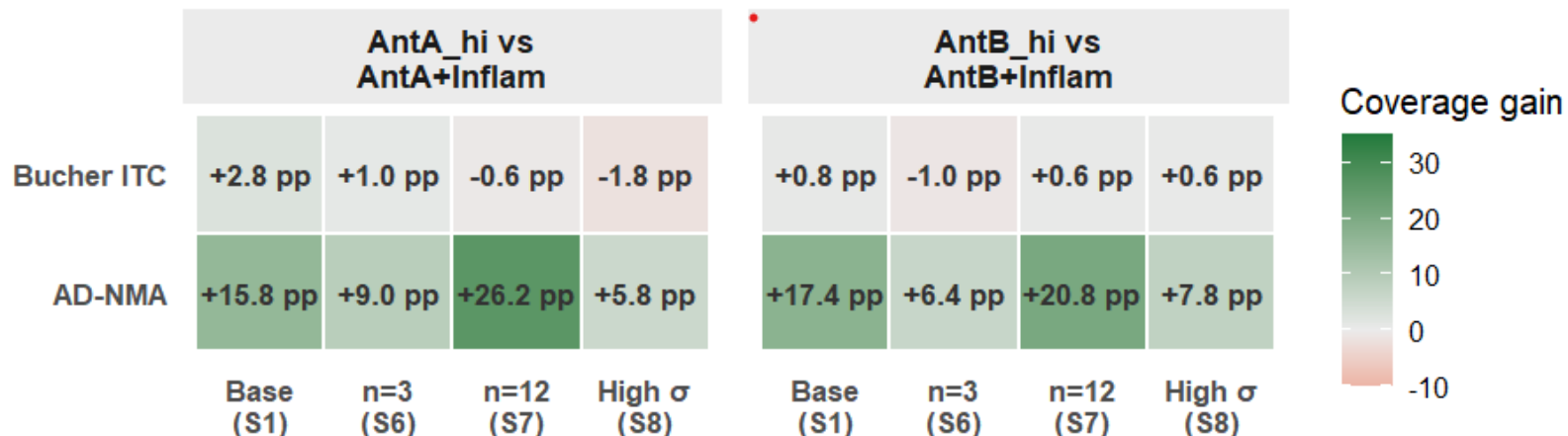
Under which network/design is the individual level data advantageous?

Positive values indicate lower RMSE with IPD-NMA



Coverage gain using IPD-NMA

Positive values in percent point (pp) indicate more coverage with IPD-NMA

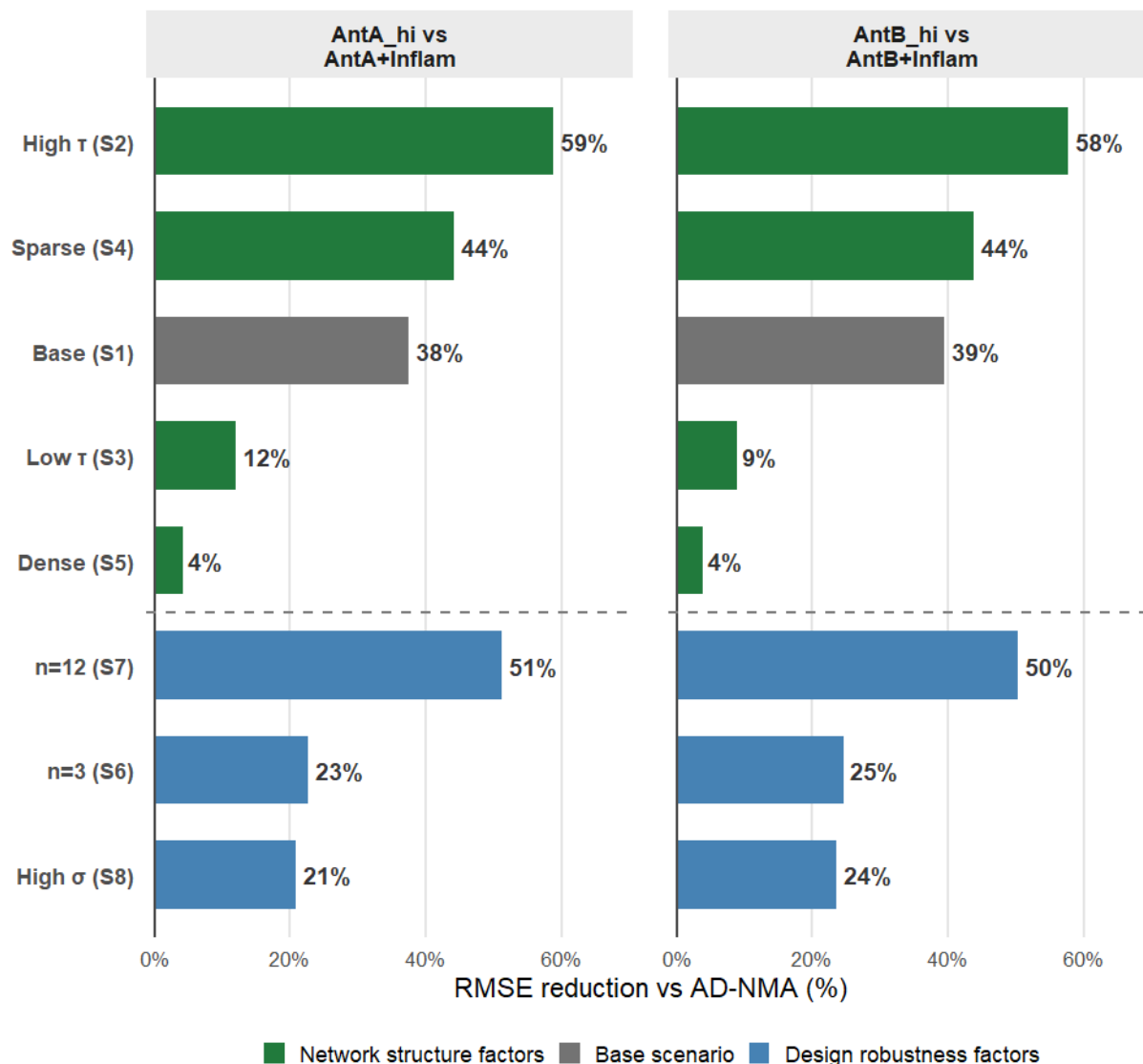


- IPD-NMA remains valuable regardless of study design.
- The advantage grows larger with higher n, showing IPD-NMA better separates donor variability from noise – distinction from AD-NMA.
- Bucher ITC was closer to IPD-NMA than AD-NMA, but IPD-NMA still showed higher precision and reliability.

IPD advantage is largest in high heterogeneity, high n, sparse settings

- *The largest IPD gain occurs when donor variability is high;*
 - as donors vary more, AD-NMA's simple averaging fails to capture that spread; leading to less accurate estimates and unreliable confidence intervals
 - when donors are similar, aggregating them loses little information.
- *Study size matters;*
 - more replicates shrink measurement noise for all methods but AD-NMA underestimates donor variability — so as other noise shrinks, this unexplained variability makes up a larger share of its error.
- *IPD-NMA advantage scales with sparsity;*
 - dense networks eliminate the advantage when every treatment appears in every donor, all methods converge.
 - when networks are sparse, unbalanced treatment-donor combinations within a study create uncanceled donor contributions to study-level means,

Drivers and robustness of the IPD-NMA advantage



- IPD-NMA gave more precise estimates vs AD-NMA in every scenario tested;
 - biggest benefit under high heterogeneity, sparse networks and with more replicates.
- The advantage held up also with fewer replicates and higher experimental noise.
- Advantages were minimal in dense networks and low heterogeneity settings.

IPD-NMA showed benefit vs AD-NMA in every scenario tested, with largest gains coming from high donor heterogeneity, more replicates, and sparse networks.

Conclusions

-  IPD-NMA provides indirect treatment estimation with higher precision and power
-  Benefit is greatest in high heterogeneity, more replicates and sparse network settings
-  IPD-NMA advantage holds up across every condition tested

For sparse preclinical networks with high heterogeneity preserving study donor-level information through IPD-NMA produces more precise and higher-powered indirect estimates than aggregate-data approaches.



Ezgi Tanriver-Ayder, Senior Principal Statistician
Oncology Discovery Statistics, GSK

Email: ezgi.x.tanriver-ayder@gsk.com

LinkedIn: [ezgitanriverayder](https://www.linkedin.com/in/ezgitanriverayder)

