



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/220040/>

Version: Published Version

---

**Article:**

Nevitt, Sarah Jane, Phillippo, David M., Hodgson, Robert et al. (2025) Application of Multilevel Network Meta-Regression in the NICE Technology Appraisal of Quizartinib for induction, consolidation and maintenance treatment of newly diagnosed FLT3-ITD-positive acute myeloid leukaemia: an External Assessment Group perspective. *PharmacoEconomics*. pp. 243-247. ISSN: 1179-2027

<https://doi.org/10.1007/s40273-024-01460-1>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial (CC BY-NC) licence. This licence allows you to remix, tweak, and build upon this work non-commercially, and any new works must also acknowledge the authors and be non-commercial. You don't have to license any derivative works on the same terms. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

**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](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



# Application of Multi-level Network Meta-Regression in the NICE Technology Appraisal of Quizartinib for Induction, Consolidation and Maintenance Treatment of Newly Diagnosed FLT3-ITD-Positive Acute Myeloid Leukaemia: An External Assessment Group Perspective

Sarah J. Nevitt<sup>1</sup> · David M. Phillippo<sup>2</sup> · Robert Hodgson<sup>1</sup> · Nicky J. Welton<sup>2</sup> · Sofia Dias<sup>1</sup>

Accepted: 21 November 2024  
© The Author(s) 2024

## 1 Introduction

Matching-adjusted indirect comparisons (MAICs) [1] are increasingly popular within National Institute for Health and Care Excellence (NICE) Single Technology Appraisals (STAs) as a method to adjust for cross-study differences in patient characteristics, which are treatment-effect modifiers [2, 3]. Matching-adjusted indirect comparisons are applicable only in a two-study indirect treatment comparison (ITC) scenario where individual participant data (IPD) are available from a study comparing treatment A versus treatment C and aggregate data (AD) from a second study comparing treatment B versus treatment C, to obtain the indirect comparison of A versus B. An inherent limitation is that MAICs provide comparative effect estimates that are applicable only to the population of the AD study and cannot be transposed to different populations [4]. Multi-level network meta-regression (ML-NMR) overcomes these limitations, allowing flexibility to generate population-adjusted ITC estimates from any number of treatments and studies, which are applicable to any specified target population [4–6].

In this commentary, we describe observations and key learnings from an External Assessment Group (EAG) perspective of the first use of ML-NMR for survival outcomes in an STA submitted to NICE for the treatment of newly diagnosed FMS-like tyrosine kinase 3 internal tandem

duplication positive (*FLT3-ITD+*) acute myeloid leukaemia (AML) [TA1013] [7].

## 2 Quizartinib for Induction, Consolidation and Maintenance Treatment of Newly Diagnosed *FLT3-ITD+* AML

NICE invited Daiichi Sankyo UK to submit evidence for the clinical and cost effectiveness of quizartinib (Vanflyta<sup>®</sup>) for the treatment of newly diagnosed *FLT3-ITD+* AML. The NICE scope outlined two comparators to be considered as part of the decision problem: standard chemotherapy (SC) and midostaurin plus SC. The midostaurin regimen represents the current standard of care in the National Health Service (NHS), so was the main comparator. Quizartinib clinical effectiveness evidence was based primarily on QuANTUM-First [8], a phase III, double-blind, randomised controlled trial of quizartinib plus SC versus placebo plus SC in patients with newly diagnosed *FLT3-ITD+* AML.

In the absence of direct comparative effectiveness evidence for quizartinib versus midostaurin, the company presented ITCs for key outcomes including overall survival. The ITCs included QuANTUM-First and the *FLT3-ITD+* subgroup of the RATIFY trial [9], a phase III, double-blind, randomised controlled trial of midostaurin plus SC versus placebo plus SC in patients with newly diagnosed *FLT3+* AML (Fig. 1).

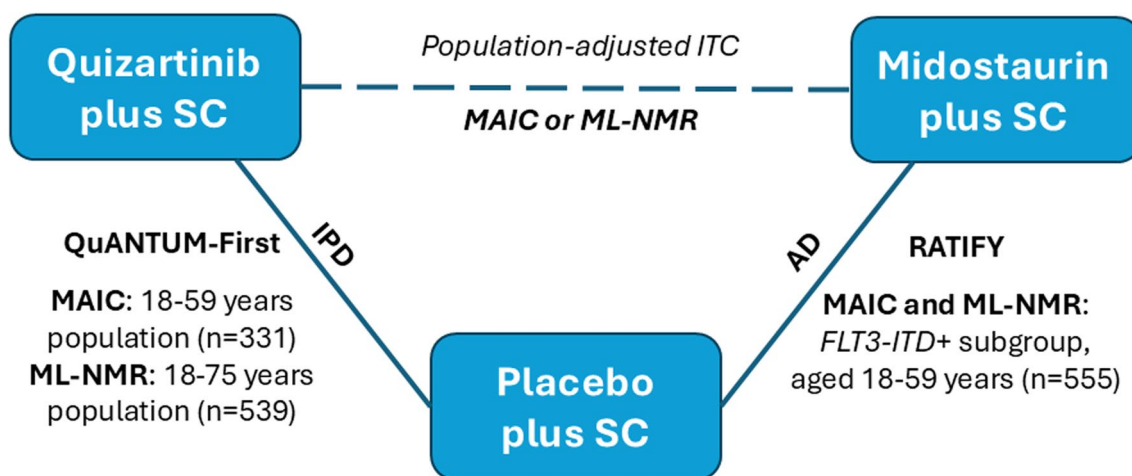
Due to differences in patient characteristics across the QuANTUM-First and RATIFY trials, the company performed MAICs that were used to inform their base-case economic analysis (Fig. 1; Table 1). Unadjusted ITCs without population adjustment were also presented for reference.

QuANTUM-First and RATIFY differed substantially in terms of design, setting and eligibility criteria. Among the

✉ Sarah J. Nevitt  
sarah.nevitt@york.ac.uk

<sup>1</sup> Centre for Reviews and Dissemination (CRD), University of York, York YO10 5DD, UK

<sup>2</sup> Bristol Technology Assessment Group, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK



**Fig. 1** Indirect treatment comparison (ITC) of the QuANTUM-First and RATIFY trials. *AD* aggregate data, *FLT3-ITD+* FMS-like tyrosine kinase 3 internal tandem duplication positive, *IPD* individual participant data, *MAIC* matching adjusted indirect comparison, *ML-NMR* multi-level network meta-regression, *SC* standard chemotherapy

**Table 1** Overall survival results for quizartinib versus midostaurin: adapted from TA1013 [7]

Included trial populations	Analysis	Target population represented	Estimand	OS: HR (95% CI) <sup>a</sup>
QuANTUM-First 18–59 years of age population and RATIFY <i>FLT3-ITD+</i> subgroup (18–59 years of age) <sup>b</sup>	MAIC <sup>c,d</sup>	‘RATIFY-like’ population	Population-average marginal	0.87 (0.56, 1.34)
	Unadjusted ITC <sup>c</sup>	Unclear <sup>f</sup>	Marginal	0.82 (0.48, 1.39)
QuANTUM-First 18–75 years of age (all ages) population and RATIFY <i>FLT3-ITD+</i> subgroup (18–59 years of age)	ML-NMR <sup>c,d</sup>	‘QuANTUM-First-like’ and ‘RATIFY-like’ population	Population-average conditional	1.02 (0.67, 1.56) <sup>g</sup>
	Unadjusted ITC <sup>e</sup>	Unclear <sup>f</sup>	Marginal	0.99 (0.68, 1.43)

*CI* confidence interval, *EAG* Evidence Assessment Group, *FLT3-ITD+* FMS-like tyrosine kinase 3 internal tandem duplication positive, *HR* hazard ratio, *IPD* individual participant data, *ITC* indirect treatment comparison, *MAIC* matching adjusted indirect comparison, *ML-NMR* multi-level network meta-regression, *NPM1* nucleophosmin 1, *OS* overall survival

<sup>a</sup>HR < 1 favours quizartinib over midostaurin

<sup>b</sup>Due to the differences in the age eligibility criteria in the two trials, in order to match the QuANTUM-First trial IPD to the RATIFY trial aggregate data, the company restricted the QuANTUM-First trial population to the population < 60 years of age (61.4% of the all ages population)

<sup>c</sup>ITC conducted by the company

<sup>d</sup>Population-adjusted ITCs were adjusted for platelet count, age, sex and *NPM1* mutation status

<sup>e</sup>ITC conducted by the EAG

<sup>f</sup>Without population adjustment, the target population of an unadjusted ITC represents a combination of the QuANTUM-First and RATIFY trial populations, assuming the populations are the same

<sup>g</sup>95% credible interval presented as ML-NMR is implemented in a Bayesian framework

most important differences were eligibility criteria relating to participants’ age. In QuANTUM-First, patients aged 18–75 years were eligible, while RATIFY was limited to patients aged 18–59 years. As is common in NICE STAs, the company only had access to IPD from their own QuANTUM-First trial. Therefore, to implement a MAIC, the company matched the QuANTUM-First trial population to the published AD of the RATIFY *FLT3-ITD+* subgroup. The resulting relative effect estimates thus reflected a ‘RATIFY-like’ population.

The EAG raised concerns, supported by clinical experts, regarding the generalisability of the RATIFY trial to the NHS population eligible to receive quizartinib. Changes in clinical practice have improved treatment outcomes since the conduct of the RATIFY trial between 2008 and 2016, and the age of the RATIFY population is younger than the current NHS population. Therefore, the EAG concluded that the results of the MAICs, reflecting a ‘RATIFY-like’ population, were not suitable for decision making. To overcome this fundamental limitation of MAICs, the EAG requested that the company also perform ML-NMR [10]

and generate results reflecting a ‘QuANTUM-First-like’ population (Table 1), which were included in a scenario economic analysis.

In the two-study scenario typically encountered in NICE STAs; data are insufficient to estimate independent effect-modifier interaction terms for each treatment. Therefore, without additional assumptions, all population-adjustment methods can only produce estimates in the AD trial population. With ML-NMR, a simplifying assumption of shared effect-modification may be made, where effect-modifier interactions are equal for quizartinib and midostaurin so that estimates can be produced for any target population [4, 5]; ML-NMR can apply this assumption for any data type, whereas for MAICs, this assumption is only applicable with continuous outcomes so could not be applied [6]. Clinical experts supported this assumption as both treatments are tyrosine kinase inhibitors targeting *FLT3* mutations. ML-NMR estimates were produced both for a ‘QuANTUM-First-like’ population and a ‘RATIFY-like’ population. The NICE committee considered that the results of the ML-MNRs were more applicable to the target NHS population and suitable for decision making [7].

In the context of decision making, it is important to note that the different methods estimate different types of population-average treatment effects. MAICs produce marginal treatment effects, whereas ML-NMR can produce either marginal or population-average conditional treatment effects. Typically these do not align where non-collapsible effect measures such as hazard ratios (HRs, for survival outcomes) and odds ratios (for dichotomous outcomes) are used [6, 11], which can lead to different decisions regarding comparative clinical or cost-effectiveness.

Marginal treatment effects are specific to a target population with given distributions of prognostic and treatment-effect modifier characteristics and, for survival outcomes, shape of the baseline hazard [6, 10]. Marginal HRs additionally depend on the length of follow-up and the censoring distribution [11]. For decision making, the marginal treatment effect represents the effect of moving everyone within the target population from treatment with midostaurin to treatment with quizartinib. The existence of covariate effects (prognostic or effect modifying) means that hazards are mathematically non-proportional at a marginal level [6, 10]; i.e. the marginal HR is time varying and a single constant HR cannot be applied.

Population-average conditional treatment effects are specific to a target population with a given distribution of treatment-effect modifier characteristics and are interpreted as the average of the individual-level treatment effects in the population, i.e. the average effect of moving each individual within the target population from treatment with midostaurin to treatment with quizartinib. Unlike marginal treatment effects, population-average conditional treatment effects do

not depend on the distribution of prognostic characteristics or the baseline hazard and may be constant over time.

Careful interpretation of population-average conditional and marginal treatment effects in the context of the decision problem is required [6]. Crucially, the different interpretations of population-average marginal and conditional effects mean that care must be taken to incorporate them appropriately into an economic model, where the primary focus is on obtaining an appropriate population-average survival curve for each treatment.

The company’s approach to modelling relative treatment effects was to apply population-average conditional HRs estimated from the ML-NMR to extrapolated survival evidence from the quizartinib arm from QuANTUM-First [7]. However, directly applying constant, population-average conditional HRs from the ML-NMR to a baseline marginal survival curve, which is unadjusted for covariates, combines incompatible conditional and marginal quantities, resulting in aggregation bias [10]. The effect of this bias on cost-effectiveness estimates is, however, unknown [7].

A correct approach to a cost-effectiveness analysis when covariate effects are present would be to use the population-average survival curves for each treatment in the target population predicted from the ML-NMR directly within the economic model. Baseline survival data for an external target population may be included within the ML-NMR model to produce coherent predictions for populations other than those of the included studies [10, 12].

However, incorporating population-average survival curves is not trivial because of the need to account for changing hazards across the model time horizon and the computational burden of accounting for uncertainty in estimates. Analysts considering the use of ML-NMR to inform economic evaluations must carefully consider how estimates should be incorporated in advance of economic model development. Similar considerations apply to other population-adjustment methods: the common approach of applying a constant marginal HR derived from a MAIC to a baseline survival curve contradicts the assumed existence of covariate effects, which are the motivation for population adjustment. Attention should also be given to how survival data are extrapolated. NICE committees and EAGs will usually expect to see the exploration of a variety of extrapolation approaches. Development of further guidance on how ML-NMR estimates can be integrated into economic analyses is an important research priority.

### 3 Key Learnings

- Despite widespread use in NICE technology appraisals, MAICs often do not generate treatment effect estimates that are generalisable to the target population.

- ML-NMR can generate effect estimates applicable to any specified target population.
- Effect estimates must be interpreted appropriately in the context of population-level decision making and applied correctly in economic modelling to avoid aggregation bias.
- Population-average survival curves for each treatment in the target population predicted from the ML-NMR should be used directly within economic models. Further practical guidance on this implementation is required.

## 4 Conclusions

This is the first application of ML-NMR for survival outcomes in a NICE technology appraisal. The NICE committee recognised the value of ML-NMR to generate treatment effect estimates in the target population and, therefore, we anticipate that the use of ML-NMR within NICE technology appraisals will increase in the future. However, further issues were raised on the appropriate interpretation of ML-NMR treatment effect estimates for decision-making, as well as the correct incorporation of results in economic models. Further guidance on the practical implementation of ML-NMR results into economic evaluations is required.

## Declarations

**Funding** This report was commissioned by the NIHR Evidence Synthesis Programme as project number NIHR131946. Please visit the HTA programme website for further project information (<https://www.nihr.ac.uk/explore-nihr/funding-programmes/health-technology-assessment.htm>). This summary of the EAG report was compiled after NICE issued the Final Guidance. The views and opinions expressed herein are those of the authors and do not necessarily reflect those of NICE or the Department of Health.

**Conflict of interest** The authors have no conflicts of interest that are directly relevant to the content of this article.

**Ethics approval** Not applicable.

**Consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Availability of data and material** The data that support this commentary are available on the NICE website (<https://www.nice.org.uk/guidance/TA1013>).

**Code availability** Not applicable.

**Author contributions** SJN and RH were jointly responsible for leading the critical review of the clinical and cost-effectiveness evidence and writing of the EAG report. DMP provided expert advice on matching adjusted indirect treatment comparisons and multi-level network meta-regression that informed the EAG report. All authors contributed to the conception and design of this article. The manuscript was initially drafted by SJN, with input from RH relating to

economic modelling content and from SD, DMP and NJW relating to methodology of population-adjusted indirect treatment comparisons. All authors had the chance to comment on draft versions of the manuscript and provided their approval for the final version to be published. SJN acts as the overall guarantor for the article.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

## References

1. Signorovitch JE, Wu EQ, Yu AP, Gerrits CM, Kantor E, Bao Y, et al. 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):935–45.
2. Phillippo D, Ades T, Dias S, Palmer S, Abrams KR, Welton N. NICE DSU technical support document 18: methods for population-adjusted indirect comparisons in submissions to NICE. Sheffield: Decision Support Unit, ScHARR, University of Sheffield; 2016.
3. Truong B, Tran LT, Le TA, Pham TT, Vo TT. Population adjusted-indirect comparisons in health technology assessment: a methodological systematic review. *Res Synth Methods*. 2023;14(5):660–70.
4. Phillippo DM, Dias S, Ades AE, Belger M, Brnabic A, Schacht A, et al. Multilevel network meta-regression for population-adjusted treatment comparisons. *J Roy Stat Soc Ser A Stat Soc*. 2020;183(3):1189–210.
5. Phillippo DM, Dias S, Ades AE, Belger M, Brnabic A, Saure D, et al. Validating the assumptions of population adjustment: application of multilevel network meta-regression to a network of treatments for plaque psoriasis. *Med Decis Making*. 2023;43(1):53–67.
6. Phillippo DM, Remiro-Azócar A, Heath A, Baio G, Dias S, Ades AE, et al. Effect modification and non-collapsibility leads to conflicting treatment decisions: a review of marginal and conditional estimands and recommendations for decision-making. 2024. Available from: <https://arxiv.org/abs/2410.11438>. Accessed 02 Dec 2024.
7. National Institute for Health and Care Excellence (NICE). Quizartinib for induction, consolidation and maintenance treatment of newly diagnosed FLT3-ITD-positive acute myeloid leukaemia. Technology appraisal guidance 2024. Available from: <https://www.nice.org.uk/guidance/TA1013>. Accessed 02 Dec 2024.
8. Erba HP, Montesinos P, Kim HJ, Patkowska E, Vrhovac R, Zak P, et al. Quizartinib plus chemotherapy in newly diagnosed patients with FLT3-internal-tandem-duplication-positive acute myeloid leukaemia (QuANTUM-First): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet*. 2023;401(10388):1571–83.

9. Rücker FG, Du L, Luck TJ, Benner A, Krzykalla J, Gathmann I, et al. Molecular landscape and prognostic impact of FLT3-ITD insertion site in acute myeloid leukemia: RATIFY study results. *Leukemia*. 2022;36(1):90–9.
10. Phillippo DM, Dias S, Ades AE, Welton NJ. Multilevel network meta-regression for general likelihoods: synthesis of individual and aggregate data with applications to survival analysis. 2024. Available from: <https://arxiv.org/abs/2401.12640>. Accessed 02 Dec 2024.
11. Daniel R, Zhang J, Farewell D. Making apples from oranges: comparing noncollapsible effect estimators and their standard errors after adjustment for different covariate sets. *Biom J*. 2021;63(3):528–57.
12. Phillippo DM. *multinma*: Bayesian network meta-analysis of individual and aggregate data. R package version 0.7.2. 2024. Available from: <https://dmphillippo.github.io/multinma/>. Accessed 02 Dec 2024.