International Journal of General Medicine

Dovepress
open access to scientific and medical research

Open Access Full Text Article

COMMENTARY

The Importance of Clinical Context and Consistency in Methodology When Using Matching-Adjusted Indirect Comparisons (MAICs) to Compare Outcomes

Katharine Batt[1,*], Robert Klamroth[2,*], Maria Elisa Mancuso[3,4,*], Andreas Tiede[5,*], Lorenzo G Mantovani[6,7,*]

[1]Department of Internal Medicine, Section of Hematology/Medical Oncology, Wake Forest University Baptist Medical Center, Winston-Salem, NC, USA; [2]Department for Internal Medicine, Vascular Medicine and Coagulation Disorders, Vivantes Hospital, Friedrichshain, Berlin, Germany; [3]Center for Thrombosis and Hemorrhagic Diseases, IRCCS Humanitas Research Hospital, Rozzano, Italy; [4]Humanitas University, Pieve Emanuele, Milan, Italy; [5]Hematology, Hemostaseology, Oncology and Stem Cell Transplantation, Hannover Medical School, Hannover, Germany; [6]Center for Public Health Research, University of Milano-Bicocca, Monza, Italy; [7]Value-Based Healthcare Unit, IRCCS MultiMedica Research Hospital, Sesto San Giovanni, Italy

*These authors contributed equally to this work

Correspondence: Katharine Batt, Wake Forest University Baptist Medical Center, 1 Medical Center Blvd, Winston-Salem, NC, 27157, USA, Tel +1 919-593-6188, Email katharine.batt@aya.yale.edu

Abstract: Hemophilia A is rare, which makes large, randomized, controlled, statistically driven, head-to-head comparison trials difficult. Matching-adjusted indirect comparisons (MAICs) are validated statistical tools designed to help make the results of non-comparative trials more comparable. The purpose of this commentary is to provide an insight into the MAIC method, in order to assist the hemophilia community with interpretation of MAIC data. It includes a comparison of the findings from previously published MAICs comparing recombinant factor replacement options and their methodologies. As MAICs are being used more often to compare treatment options for patients with hemophilia A, it is paramount that robust and consistent methodologies for cross-trial comparisons are used and that all efficacy analysis findings are linked to factor utilization.

Keywords: hemophilia, MAIC, methodology

Hemophilia A is a rare blood clotting disorder, affecting around 1 in 5000 male births. This means study populations are often small, which makes large, randomized, controlled, statistically driven, head-to-head comparison trials difficult.[1,2] Matching-adjusted indirect comparisons (MAICs) are validated statistical tools designed to help make the results of non-comparative trials more comparable. They compare therapeutic products by combining and re-weighting known individual patient data from one clinical trial with the published baseline summary statistics of comparator treatments as derived from reference clinical trials.[3,4] The value of MAICs in rare disease therapeutic areas such as hemophilia has been acknowledged by various health assessment bodies, leading to an increase in the number of MAICs reported in the last few years.[1] The limitations of MAICs and of the methodologies used in these types of analysis are less widely understood, which can negatively impact data interpretation and application to clinical practice. The purpose of this commentary is to provide an insight into the MAIC method, in order to assist the hemophilia community with interpretation of MAIC data. The MAIC methodology has been published previously, and the National Institute for Health and Care Excellence (NICE) guidelines provide a reputable overview of the approach.[1,3,5,6] This method is also explained in an animation (https://doi.org/10.6084/m9.figshare.18705461) comparing data on the extended-half-life factor VIII replacement damoctocog alfa pegol from the PROTECT VIII trial with published aggregate data from trials of efmoroctocog alfa, rurioctocog alfa pegol, antihemophilic factor (recombinant), plasma/albumin-free method (octocog alfa) and turoctocog alfa pegol.

Received: 22 February 2024
Accepted: 14 July 2024
Published: 7 September 2024

There are a number of core steps required for comparisons made between individual patient data (IPD) from trials with aggregate data to be valid.[3] As a first step, the criteria used in trial selection must be transparent and methodological. Often, this is accomplished via a systematic literature review that clearly stipulates trial characteristics, including study designs and patient inclusion/exclusion criteria; factors that represent key variations between trials and clarify the presence or exclusion of certain treatment arms. There are many reasons why a trial may be selected, but data from a particular arm may not be appropriate for inclusion in the MAIC. For example, a comparison may exclude an arm that examines a dosing regimen that is not in the approved labelling for both therapeutic products under evaluation. Or it may exclude patients assigned to a specific arm with a positive selection step, such as a lead-in period, during which patients have an opportunity to change dosing or switch to another regimen that would significantly bias outcomes (self-selection for worse disease). Though MAICs are a powerful tool for comparing studies, the methodology cannot compensate for factors such as these. As decisions regarding the inclusion or exclusion of particular arms or patient groups are likely to be subjective, it is important that the reasoning is clearly communicated to the readers so that they may make an informed interpretation of the findings and their equivalence. Once the overall study design has been analyzed and decisions made regarding the comparability of other trial designs, the baseline characteristics for the chosen patient populations, such as specific disease-related characteristics like bleed history and joint health, as well as other prognostic variables and sources of heterogeneity, should be analyzed. In the event that two trials have different baseline patient populations, the patients can be compared through a process of adjusting the baseline populations and matching like with like. Even if characteristics are reported in different formats, for example annualized versus monthly utilization data, these differences can help to clarify the differences between the two populations. In deciding which characteristics should be compared to highlight the most important data across trials, expert opinion (and/or overview of previously published MAICs) is relied upon. For example, were a trial to not report anything about the treated population's baseline hemoglobin, we could not use this characteristic in a MAIC, but it does not mean that it is not relevant; any reader would struggle to trust the results of the comparison, because we could not safely say that the populations were ever comparable at all. After identifying any sources of inter-trial variation, an assessment should be performed on the availability of IPD. These data will allow cross-trial differences to be offset. The next step is to identify and match the definitions of the outcome measures, while considering the statistics used and the clinical relevance of dosages and utilization data. Plainly, improved clinical outcomes can be achieved with certain doses at certain frequencies, and this should be considered as part of a comparison. The use of sensitivity analyses is required when exact matching of the definitions of outcome measures is not possible; recalculating outcomes under alternate assumptions about the inputs can indicate where meaningful relationships exist. Finally, the trial populations should be matched; at this step, any patient in the IPD group who would not meet entry requirements for the comparator trial should be excluded and all remaining patients assigned a weighted score. This weighted score matches their baseline characteristics to those reported for the comparator trial. It should be noted that a fair comparison cannot be made when there is an inability to match inclusion/exclusion criteria or outcome measures. Cross-trial differences that have not been accounted for and unnecessary exclusion of patients can introduce bias. Only a randomized, controlled head-to-head trial can offer reassurance that the results are not biased by unaccounted for differences. In addition, to enable the matching of all baseline characteristics, MAICs require there to be more patients than the number of baseline characteristics. While common comparator arms, for example two trials each using octocog alfa as a comparator arm, are not mandatory, the presence of a common comparator arm facilitates the validation of matching.[1,3] As a consequence of the matching process, a MAIC reflects potential outcomes in a population similar to the one in the comparator trial. When interpreting the results, generalizability to different patient populations should be assessed. Finally, and as mentioned earlier, some trials simply cannot be compared in a MAIC due to irreconcilable differences between them.

A Comparison of the Findings from Previously Published MAICs Comparing Recombinant Factor Replacement Options and Their Methodologies

In 2019, Batt et al reported the results of a MAIC showing similar bleeding outcomes and a 20–40% lower factor utilization when comparing damoctocog alfa pegol results from the Phase 2/3, partially randomized PROTECT VIII trial

with published aggregate data for efmoroctocog alfa, rurioctocog alfa pegol and octocog alfa.[1] Post-match data for each comparator trial are shown in Table 1. To ensure comparability between patient populations, key inclusion and exclusion criteria for the comparator trials were applied to the BAY 94–9027 population. The variability in treatment frequency across comparator treatments was accommodated using data from the three different prophylactic dosing regimens (twice weekly, every 5 days and every 7 days) in PROTECT VIII. For efmoroctocog alfa, individualized and weekly prophylaxis arms from the ALONG trial were pooled, along with a sensitivity analysis using only data from the individualized efmoroctocog alfa treatment arm. Separate analyses were conducted for the data from two published octocog alfa trials, 2004 (standard prophylaxis arm) and 2012 (pooled standard and pharmacokinetics-tailored prophylaxis arms), due to differences in the calculation of the annualized bleeding rates (ABR). For rurioctocog alfa pegol, the prophylaxis arm in the PROLONG-ATE trial was used. A logistics regression model was then used to estimate individual patient weightings, and method of moments to estimate the parameters in this model. The use of categorical-based variables allowed for a high number of these to be matched across treatments, and all summary statistics for all baseline characteristics were exactly balanced after matching.[1]

Similar results were reported by Vashi et al in 2021, where a 26% lower factor utilization was demonstrated for damoctocog alfa pegol compared with turoctocog alfa pegol, while maintaining similar bleeding outcomes.[7] In this MAIC, IPD from the prophylaxis arms in the PROTECT VIII main study were compared to aggregate data from the prophylaxis arm of the partially randomized PATHFINDER 2 main study; the method for unanchored trials was applied due to the non-random allocation of patients to the on-demand regimens in each trial, resulting in the outcomes observed being non-comparable to those observed in the prophylaxis arms. Again, as many variables as possible were matched, and both trials had similar inclusion criteria and similar outcome definitions.[7] The reporting of these trials is consistent with that of Bonanad et al in 2021, where both efficacy and utilization of lonoctocog alfa were compared with those of efmoroctocog alfa and octocog alfa.[8] In contrast, in 2021, Hakimi et al reported efficacy findings from a MAIC without linking these to any utilization data.[9] The authors reported a superior efficacy profile for efmoroctocog alfa in the A-LONG study compared with damoctocog alfa pegol in the PROTECT VIII study following MAIC analysis.[9] The discrepancies between the reported results from different research teams may arise from the variation in methodologies used. Batt et al excluded IPD from one patient that did not meet the inclusion criteria for A-LONG, whereas, in the Hakimi et al study, A-LONG IPD from the weekly 65 IU/Kg efmoroctocog alfa prophylaxis group were omitted and only the IPD from the efmoroctocog alfa prophylaxis group receiving 25–65 IU/kg every 3–5 days were included. In contrast, the aggregate data used from PROTECT VIII included patients receiving damoctocog alfa pegol prophylaxis 30–40 IU/kg twice-weekly, 45–60 IU/kg every 5 days, and 60 IU/kg every 7 days.[1,9] This difference likely reflects a difference in illness severity in the groups included and may make the results harder to interpret. We hope to support the hemophilia community in gaining a better understanding of the MAIC method, allowing them to identify differences in methodology – including the absence of measures that provide clinical context – and to better interpret the results. The use of categorical variables allows for a greater number of variables to be matched compared with the use of numerical variables, for example "severe hemophilia" versus proportion of normal clotting activity, because categories have the same definitions whereas numerical values are more likely to be different. In addition, reporting of efficacy profiles linked to utilization data gives clinical context to the comparison data. Consistency in methodology across all MAICs is key to achieving a fair comparison. Methodological differences can lead to selection bias, for example through omitting patient subgroups; elimination of clinical context, as a result of excluding factor utilization; and results being invalidated when outcomes are not matched, including comparisons of ABR, target bleeds and factor utilization.[3] As MAICs are being used more often to compare treatment options for patients with hemophilia A, it is paramount that robust and consistent methodologies for cross-trial comparisons are used and that all efficacy analysis findings are linked to factor utilization. Despite the limitations of the MAIC methodology, such as the incomparability of some studies and the potential bias introduced by unaccounted for cross-trial differences, MAICs may then aid the hemophilia community in making informed choices when selecting factor replacements.

Table 1 Compilation of Post-Match Bleeding Outcomes and Utilization Data from MAICs with FVIII Replacements

Batt et al 2019[1,a]										
	Damoctocog alfa pegol[b] 30–40 IU/kg 2×W, 45–60 IU/kg E5D, 60 IU/kg E7D	**Efmoroctocog alfa[b] 25–65 IU/kg E3–5D, 65 IU/kg weekly**	**Damoctocog alfa pegol[b] 30–40 IU/kg 2×W, 45–60 IU/kg E5D, 60 IU/kg E7D**	**Efmoroctocog alfa[b] 25–65 IU/kg E3–5D (sensitivity analysis)**	**Damoctocog alfa pegol 30–40 IU/kg 2×W, 45–60 IU/kg E5D, 60 IU/kg E7D**	**Rurioctocog alfa pegol 45 ±5 IU/kg 2×W**	**Damoctocog alfa pegol 30–40 IU/kg 2×W, 45–60 IU/kg E5D, 60 IU/kg E7D**	**Octocog alfa (2004) 25–40 IU/kg 3×W/EOD**	**Damoctocog alfa pegol 30–40 IU/kg 2×W, 45–60 IU/kg E5D, 60 IU/kg E7D**	**Octocog alfa (2012) 20–40 IU/kg every 48±6 h, 20–80 IU/kg every 72±6 h**
Mean ABR	3.77	3.90	4.25	2.91	3.95	3.70	4.28	6.30	1.87	1.80
Zero bleeds (%)	34.1	40.7	35.2	45.3	38.9	39.6	41.5	29.9	40.0	33.3
Factor utilization (IU/kg/W)	66.5	82.2*	66.4	85.4**	64.3	87.4	64.0	107.5	63.6	109.9
Factor utilization (IU/kg/Y)	3458.0	4274.4	3452.8	4440.8	3343.6	4544.8	3328.0	5590.0	3307.2	5714.8
Vashi et al 2021[7]										
	Damoctocog alfa pegol[b] 30–40 IU/kg 2×W, 45–60 IU/kg E5D, 60 IU/kg E7D	**Turoctocog alfa pegol[b] 50 IU/kg E4D**								
Mean ABR	4.10	3.70								
Zero bleeds (%)	0.41	0.40								
Factor utilization (IU/kg/W)	68.3	93.2								
Factor utilization (IU/kg/Y)	3552.43	4845.00***								
Bonanad et al 2021[8]										
	Lonoctocog alfa[b] 20–50 IU/kg 2–3×W/EOD	**Efmoroctocog alfa[b] 25–65 IU/kg E3–5D**	**Lonoctocog alfa 20–50 IU/kg 2–3×W/EOD**	**Octocog alfa (2004)[c] 25–40 IU/kg 3×W/EOD**	**Lonoctocog alfa 20–50 IU/kg 2–3×W/EOD**	**Octocog alfa (2012) 20–40 IU/kg every 48±6 h**				
Mean ABR	3.44	2.91	4.64	6.30	1.61	1.60				
Zero bleeds (%)	39.4	45.3	36.6	29.9	41.8	40.6				
Factor utilization (IU/kg/W)	85.5	89.1	NR	NR	81.9	110.9				
Factor utilization (IU/kg/Y)	4444.0	4632.0	NR	NR	4261.0	5768.0*				

Hakimi et al 2021[9,d]										
	Damoctocog alfa pegol 30–40 IU/kg 2×W, 45–60 IU/kg E5D, 60 IU/kg E7D	**Efmoroctocog alfa 25–65 IU/kg E3–5D**								
Mean ABR	4.9	3.0								
Zero bleeds (%)	38.2****	46.5								
Factor utilization (IU/kg/W)	NR	NR								
Factor utilization (IU/kg/Y)	NR	NR								

Notes: *$P < 0.0001$; **$P < 0.001$; ***$P < 0.005$; ****MD – 1.9 (95% CI: −3.5 to −0.4). [a]Standard error data were unavailable for rurioctocog alfa pegol and octocog alfa. Therefore, no *P*-values were calculated for these comparisons. [b]Indicates that reported factor utilization is mean data; all other factor utilization data are median factor utilization. [c]Data available were not sufficient to allow comparison of annualized factor utilization. [d]Weekly 65 IU/Kg efmoroctocog alfa prophylaxis group was omitted from the analysis. 2×W, twice weekly; ABR, annualized bleeding rate; CI, confidence interval; E3–5D, every 3–5 days; E4D, every 4 days; E5D, every 5 days; E7D, every 7 days; EOD, every other day; h, hour; MD, mean difference; NR, not reported; W, week; Y, year.

https://doi.org/10.2147/IJGM.S464226 DovePress

Abbreviations

ABR, annualized bleeding rate; IPD, individual patient data; MAIC, matching-adjusted indirect comparison.

Acknowledgments

The authors thank Graeme Baldwin and Lianne Holloway, of Darwin Healthcare Communications (Oxford, England), for providing medical writing support, which was fully funded by Bayer, in accordance with Good Publication Practice (GPP) guidelines.

Disclosure

KB: Consultancy services to Takeda Precisionheor; Cogent Biosciences, FTI Consulting, EntityRisk; Stock owned in Merck, Sanofi and J&J. RK: Honoraria, research grants or advisory boards: Bayer, BioMarin, Biotest, CSL Behring, Grifols, NovoNordisk, Octapharma, Pfizer, Roche/Chugai, Sanofi, SOBI, Takeda. MEM: Research grants from Bayer, CSL Behring, Novo Nordisk and Takeda and has acted as paid speaker/consultant/advisor for Bayer, BioMarin, CSL Behring, Grifols, Kedrion, LFB, Novo Nordisk, Octapharma, Pfizer, Roche, Sanofi, Sobi, Spark Therapeutics, Takeda and UniQure. AT: Grants for research / study support: Bayer, BioMarin, Biotest, Chugai, CSL Behring, Novo Nordisk, Octapharma, Pfizer, Roche, SOBI, Takeda. Honoraria for lectures or consultancy: Bayer, BioMarin, Biotest, Chugai, CSL Behring, Novo Nordisk, Octapharma, Pfizer, Roche, SOBI, Takeda. LGM: Grants and personal fees from Bayer AG, Daiichi Sankyo, Pfizer and Boehringer Ingelheim. The authors report no other conflicts of interest in this work.

References

1. Batt K, Gao W, Ayyagari R, et al. Matching-adjusted indirect comparisons of annualized bleeding rate and utilization of BAY 94-9027 versus three recombinant factor VIII agents for prophylaxis in patients with severe hemophilia A. *J Blood Med*. 2019;10:147–159. doi:10.2147/JBM.S206806
2. Nursing Hemophilia blog. Hemophilia incidence and statistics. Available from: https://nursinghemophilia.wordpress.com/incidence-and-statistics. Accessed August 15, 2022.
3. Signorovitch JE, Sikirica V, Erder MH, et al. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. *Value Health*. 2012;15(6):940–947. doi:10.1016/j.jval.2012.05.004
4. Thom H, Jugl SM, Palaka E, Jawla S. Matching adjusted indirect comparisons to assess comparative effectiveness of therapies: usage in scientific literature and health technology appraisals. *Value Health*. 2016;19(3):A100–A101. doi:10.1016/j.jval.2016.03.1723
5. Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. NICE DSU Technical Support Document 18: methods for population-adjusted indirect comparisons. NICE; 2016. Available from: https://www.sheffield.ac.uk/media/34216/download. Accessed August 15, 2022.
6. Signorovitch JE, Wu EQ, Yu AP, 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):935–945. doi:10.2165/11538370-000000000-00000
7. Vashi P, Batt K, Klamroth R, et al. Indirect treatment comparison of damoctocog alfa pegol versus turoctocog alfa pegol as prophylactic treatment in patients with hemophilia A. *J Blood Med*. 2021;12:935–943. doi:10.2147/JBM.S321288
8. Bonanad S, Núñez R, Poveda JL, et al. Matching-adjusted indirect comparison of efficacy and consumption of rVIII-SingleChain versus two recombinant FVIII products used for prophylactic treatment of adults/adolescents with severe haemophilia A. *Adv Ther*. 2021;38(9):4872–4884. doi:10.1007/s12325-021-01853-0
9. Hakimi Z, Santagostino E, Postma MJ, Nazir J. Recombinant FVIIIFc versus BAY 94-9027 for treatment of patients with haemophilia A: comparative efficacy using a matching adjusted indirect comparison. *Adv Ther*. 2021;38(2):1263–1274. doi:10.1007/s12325-020-01599-1