PO28

Does heterogeneity in patient population, study design, and endpoint definitions affect indirect comparisons of acute migraine treatments? A methodological assessment

Richard B Lipton[1], Phoebe Cowley[2], Emma Benbow[3], Tim Reason[4], Anthony J Hatswell[5], Aaron Jenkins[6], Lucy Abraham[7], Phillip Saccone[8], Haitao Chu[9], Terry Fullerton[10], and Cristina Tassorelli[11]

[1]Department of Neurology, Albert Einstein College of Medicine, Bronx, NY, USA; [2]Petauri Evidence; [3]Estima Scientific; [4]Estima Scientific; [5]Petauri Evidence; [6]Pfizer; [7]Pfizer; [8]Pfizer; [9]Pfizer; [10]Pfizer; [11]Department of Brain and Behavioural Science, University of Pavia, Pavia, Italy

INTRODUCTION

The lack of direct comparative trials in the field of migraine has prompted an increasing number of indirect comparisons, including most recently a systematic review and meta-analysis of acute migraine treatments by Karlsson et al., arguably the most comprehensive identifying over 100 studies for a large number of therapeutic options.[1]

The studies identified included triptans, but also contemporary treatments such as gepants. The co-primary efficacy endpoints explored were pain freedom at 2 hours and sustained pain freedom from 2-24 hours. There were seven secondary endpoints, including pain relief at 2 hours, pain relapse within 2-24 hours and use of rescue medication within 2-24 hours.

When looking across studies, indirect treatment comparisons (ITCs) can be used to provide comparative estimates using Network Meta Analysis (NMA) provided the necessary assumptions are met, else results would be undermined. A key assumption in NMA being exchangeability. This principle assumes that all included trials are sufficiently similar, meaning they measure the same outcomes in comparable patient groups, without systematic differences or patterns that could bias the results.

Given both the large number of studies, and that studies were conducted over a 30-year period, we used the work of Karlsson et al. as a case-based exercise to identify and understand any sources of heterogeneity in the included studies that may undermine the assumption of exchangeability. Where identified, we investigated how these might affect ITCs results using both 'standard' NMA methods and more sophisticated approaches.

METHODS

- **Initial review** – In order to assess heterogeneity, a critical review of previously performed ITCs in acute migraine was carried out.
  - The most commonly used ITCs were some form of meta-analysis, generally NMA.[2-4]
  - Heterogeneity is generally commented on but not formally addressed.
  - Publications that did explore heterogeneity typically used a "random effects" model when there was an I^2 value >50%, with no adjustments made for lower values.[5-7]

Based on review of previously published ITCs, studies identified by Karlsson et al. were investigated, with several issues identified.

- The intention to treat (ITT) population was used in the NMA, rather than the mITT/treated/safety population. As not all patients *intended* to be treated had an eligible migraine attack in the studies, results would be affected by classing these patients as non-responders.
- All licensed doses of treatments were pooled, implying no dose-response relationship exists – in contrast to the primary studies.
- Unlicensed intervention arms were excluded, potentially omitting [informative] indirect links.

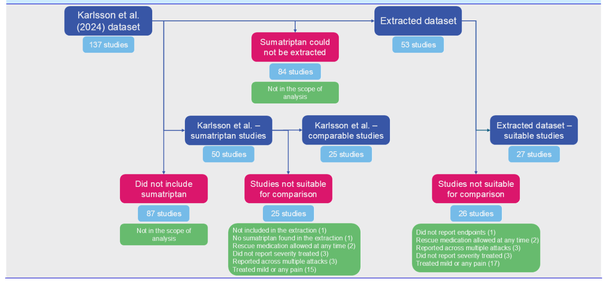
- **Re-extraction** – Based on the initial review, exploratory analysis data for sumatriptan studies were re-extracted. Sumatriptan was selected as it was the most frequently studied intervention, included in trials over a long period – from registrational studies in the 90s, to being an active comparator in contemporary studies.
- **Re-analysis** – NMA was then performed using the *netmeta* package in R (as in Karlsson et al), with a sequential approach taken to investigate the impact of heterogeneity.
  - Using the treatment population over the ITT.
  - The same approach as Karlsson et al. were run on subsets of the original and re-extracted data (**Figure 1**).
  - The endpoints identified as being most important were pain freedom, pain relief and sustained pain freedom.
- **Areas of heterogeneity** – Although the re-extraction resolved a number previously identified issues, further sources of heterogeneity were identified. The key sources of heterogeneity are shown in **Table 1**, along with any actions taken in subsequent analyses.
- **De novo analyses** – More sophisticated methods (Bayesian hierarchical models, BHM) were applied to account for multi-level data, using the *brms* R package.
  - Implementing a "class effect" ensuring drug doses were linked.
  - All timepoints used for endpoints were explored (as opposed to only 2 hours).

Table 1. Areas of heterogeneity identified in sumatriptan studies

	Details	Approach for inclusion
Population	i. Karlsson et al. used the ITT population in their NMA, however not all patients had an eligible migraine attack, and so received no treatment.	i. The use of only 'exposed' patients i.e., those that had a migraine in the study period.
	ii. Populations included in contemporary studies included patients who have previous triptan failures, whereas earlier trials had more (or exclusively) triptan naïve patients.	ii. Previous triptan use was poorly reported across studies, this therefore will remain a limitation of any analyses.
	iii. Over time, the inclusion and exclusion criteria in studies changed, mainly relating to cardiovascular health as an exclusion criteria in triptan trials.	iii. Without access to patient level data from studies, this difference is not able to be accounted for in analyses.
	iv. The majority of studies included only moderate/severe migraines, however some studies included patients with 'mild' attacks – either exclusively, or as well as moderate to severe migraine.	iv. Only studies reporting effect sizes in moderate to severe pain were included.
Intervention	i. Karlsson et al. pooled the effects of all licensed doses of a drug.	i. Full pooling assumes identical effects and omits any dose-response relationship. Hierarchical models are required to account for the "class effect" of a drug but allowing doses to vary around the central estimate.
Endpoints	i. Early studies used paper diaries, with a move to the use of e-diaries in later studies; potentially impacting endpoints given the prevalence of photophobia.	i. Too few studies are available to reliably estimate the impact of e-diaries, with no action possible.
	ii. Endpoints were measured at different timepoints across studies.	ii. More sophisticated models are able to account for the relationship between the various timepoints.
	iii. Some studies report endpoints at multiple timepoints but, as per designated primary endpoints, only single timepoints were included.	iii. Hierarchical models are able to account for multiple observations which are clustered at the study level.
	iv. Some studies did not use the same endpoints, but had results estimated using proxy endpoints – for example as using lack of sustained effect rather than sustained response.	iv. Studies that did not report the same endpoints were removed from the dataset rather than the results extrapolated.
Study Design	i. Crossover studies had primary endpoints pooled across sequential treatments. These studies have different effects to parallel group studies.	i. Crossover studies that did not report the first attack of a parallel stage (and thus were comparable to parallel group studies) were removed.
	ii. Multiple attack studies had primary endpoints based over multiple attacks, whereas the majority of studies were based on the first attack.	ii. Studies that reported endpoints across attacks were excluded unless they reported the first attack specifically.

Abbreviations: ITT, intention to treat; NMA, network meta-analysis.

Figure 1. Flowchart of included studies



RESULTS

- In total, twelve additional models (4 per endpoint) were run to assess the impact of removing studies that were not comparable and another two were run to see the effect of having all timepoints available.
- For pain freedom and relief at 2 hours, the three NMA models using the Karlsson extracted data showed an overall decrease in heterogeneity (assessed using I^2), whereas sustained pain freedom showed no real pattern in heterogeneity from the removal of studies.
- The NMAs using the extracted data were compared to the 'Comparable' Karlsson data and showed a decrease in I^2 for sustained pain freedom from 2-24 hours but an increase in pain freedom and pain relief at 2 hours. I^2 values from all NMA models are shown below (**Table 2**).

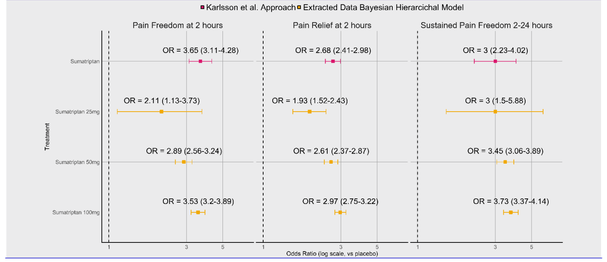
Table 2. Heterogeneity in sequential NMA models

	Karlsson estimate	Only sumatriptan studies	Only comparable sumatriptan studies	Only re-extracted comparable sumatriptan studies
Pain freedom at 2 hours	51.0%	45.1%	36.6%	39.56%
Pain relief at 2 hours	44.6%	46.0%	33.2%	39.69%
Sustained pain freedom 2-24 hours	46.8%	34.6%	53.7%	45.17%

Abbreviations: NMA, network meta-analysis.

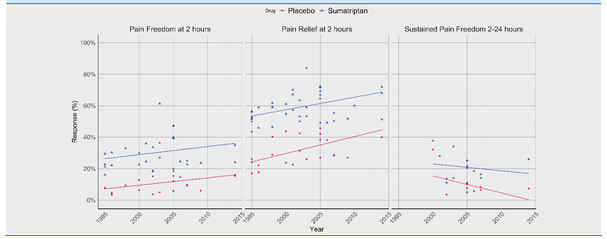
- When the NMAs were compared to the BHM, there was variation in the sumatriptan odds ratio vs placebo (OR) seen across models, as anticipated given the more homogenous studies included.
- There was also a difference between the ORs from different doses of sumatriptan vs placebo in the BHM, for example pain freedom at 2 hours model having a range in OR of 1.42 between the three doses (2.11 for 25mg to 3.53 for 100mg, **Figure 2**).
- As well as the dose impacting estimates, point estimates in the BHM were different to those in the original Karlsson et al. analysis, due to the exclusion of non-comparable studies, appropriate data use, and inclusion of all timepoints.

Figure 2. Forest plot of results from Karlsson extracted sumatriptan comparable studies for NMA and Extracted data Bayesian Hierarchical Model



- Using the re-extracted data, trends across years were plotted, with apparent increasing absolute efficacy for both placebo and sumatriptan in pain freedom and pain relief (**Figure 3**). This finding is not currently accounted for in analyses and further undermines exchangeability, and likely affects relative effectiveness estimates.

Figure 3. Absolute response of sumatriptan and placebo over time for pain freedom , pain relief and sustained pain freedom



CONCLUSIONS

- Using the same methods as a published NMA, restriction to only sumatriptan studies that could be classed as exchangeable reduced the heterogeneity seen.
- A more sophisticated analytical approach (a BHM) allowed for the inclusion of multiple doses and timepoints, affecting point estimates of efficacy.
- Even using appropriate methods, inconsistency across studies remains, most notably an apparent temporal trend towards increasing responses. If a 'true' pattern, exchangeability would be violated, impacting any resulting analyses.
- Acute migraine studies, at times conducted decades apart, show high levels of heterogeneity which would undermine a naïve NMA and any decision making based on such analysis.

REFERENCES

1: Karlsson, W.K. *et al.* (2024) *BMJ*, p. e080107. 2: Li, G. *et al.* (2023). *The Journal of Headache and Pain*, 24(1), p. 129. 3: Yang, C.-P. *et al.* (2021). *JAMA Network Open*, 4(10), p. e2128544. 4: Xu, H. *et al.* (2016). *The Journal of Headache and Pain*, 17(1), p. 113. 5: Song, Z. *et al.* (2025). *Pain and Therapy*, 14(2), pp. 639–653. 6: Lee, S. *et al.* (2022). *European Journal of Clinical Pharmacology*, 78(9), pp. 1365–1376. 7: Ha, D.K. *et al.* (2021). *Clinical Drug Investigation*, 41(2), pp. 119–132.

CONFLICTS OF INTEREST

Phoebe Cowley and Anthony J Hatswell are employees of Petauri Evidence who were engaged by Pfizer to conduct the analyses. Emma Benbow and Tim Reason are employees of Estima Scientific who were engaged by Pfizer to conduct the analyses. Richard B Lipton is on an Editorial board: Neurology, Cephalalgia (not paid); Senior advisor: Headache (not paid); research support: NIH, FDA, National Headache Foundation, Marx Foundation; research grants: Allergan/AbbVie, Amgen, Novartis; Reviewer: NIA, NINDS; consultant/advisory board/honoraria: AEON, Allergan/AbbVie, Amgen, Biodelivery Sciences International Biohaven, CoolTech LLC, Dr Reddy's Laboratories, electroCore, Eli Lilly, GlaxoSmithKline, Impel Neuropharma, Lundbeck , Manistee, Merck, Novartis, Satsuma, Teva, Vedanta; Royalties from Wolff's Headache (8th ed), Informa; Stock/stock options: Axon, Biohaven, CoolTech, and Manistee.

Cristina Tassorelli received honoraria for participation in clinical trials, contribution to advisory boards or oral presentations from: Allergan, Electrocore, Eli-Lilly, Novartis and Teva. Headache research at the IRCCS Mondino Foundation in Pavia is supported by the Italian Ministry of Health and the European Union. CT has no ownership interest and does not own stocks of any pharmaceutical company. CT serves as Chief Section Editor of Frontiers in Neurology – Section Headache Medicine and Facial Pain and on the editorial board of Journal of Headache and Pain. CT is member of the Clinical Trials Committee of the International Headache Society (IHS). Lucy Abraham, Haitao Chu, Aaron Jenkins and Phillip Saccone are employees and stockholders of Pfizer. Terry Fullerton previously worked for Pfizer.

ACKNOWLEDGEMENTS

This study was funded by Pfizer Inc. Medical writing support was provided by Phoebe Cowley and Anthony J Hatswell of Petauri Evidence.

Electronic Poster: Please scan this Quick Response (QR) code with your smartphone app to view an electronic version of this poster. If you do not have access to a smartphone, please access the poster via the following link: https://scientificpubs.congressposter.com/p/fcmle864i5cfkkt2

Presented at the European Headache Congress, December 3–6, 2025, Lisbon, Portugal