Effectiveness of multiple disease-modifying therapies in relapsing-remitting multiple sclerosis: causal inference to emulate a multiarm randomised trial

Ibrahima Diouf,1,2 Charles B Malpas 3,1,3 Sifat Sharmin 4,4, Izanne Roos 4, Dana Horakova,4,5 Eva Kubala Havrdova,4,5 Francesco Patti,6 Vahid Shaygannejad,7 Serkan Ozakbas,8 Sara Eichau 9,5 Marco Onofri 4,10 Alessandra Lugareso 4,11,12 Raed Alroughani,13 Alexandre Prat,14,15 Pierre Duquette,14,15 Murat Terzi,14,15 Cavit Boz,16,17 Francois Grand’Maison,18 Patrizia Sola,19 Diana Ferraro 19,19 Pierre Grammond,20 Bassem Yamout,21,22 Ayse Altintas,23,24 Oliver Gerlach,25,26 Jeannette Lechner-Scott,27,28 Roberto Bergamaschi,29 Rana Karabudak,30 Gerardo Iuliano,31 Christopher McGuigan,32 Elisabetta Cartechni,33 Stella Hughes,34 Maria Jose Sa,35 Claudio Solaro 36,37 Ludwig Kappos 38,39 Suzanne Hodgkinson,40 Mark Slee,41 Franco Granella,42 Koen de Gans,43 Pamela A McCombe 44,47 Radek Ampapa,49 Anneke van der Walt 40,46,47 Helmut Butzkueven,46,47 José Luis Sánchez-Menoyo,48 Steve Vucic 49, Guy Laureys,50 Youssef Sidhom 51,52 Riadh Gouider 53,52 Tamara Castillo-Trivino 53,54 Orla Gray,54 Eduardo Aguera-Morales,55 Abdullah Al-Asmi,56,57 Cameron Shaw,58 Talal M Al-Harbi,59 Tunde Csepany,60 Angel P Sempere,61 Irene Treviño Frenk 62,63 Elizabeth A Stuart,63 Tomas Kalincik 63,1,3

ABSTRACT

Background Simultaneous comparisons of multiple disease-modifying therapies for relapsing-remitting multiple sclerosis (RRMS) over an extended follow-up are lacking. Here we emulate a randomised trial simultaneously comparing the effectiveness of six commonly used therapies over 5 years.

Methods Data from 74 centres in 35 countries were sourced from MSBase. For each patient, the first eligible intervention was analysed, censoring at change/discontinuation of treatment. The compared interventions included natalizumab, fingolimod, dimethyl fumarate, teriflunomide, interferon beta, glatiramer acetate and no treatment. Marginal structural Cox models (MSMs) were used to estimate the average treatment effects (ATEs) and the average treatment effects among the treated (ATT), rebalancing the compared groups at 6-monthly intervals on age, sex, birth-year, pregnancy status, treatment, relapses, disease duration, disability and disease course. The outcomes analysed were incidence of relapses, 12-month confirmed disability worsening and improvement.

Results 23,266 eligible patients were diagnosed with RRMS or clinically isolated syndrome. Compared with glatiramer acetate (reference), several therapies showed a superior ATE in reducing relapses: natalizumab (HR=0.44, 95% CI=0.40 to 0.50), fingolimod (HR=0.60, 95% CI=0.54 to 0.66) and dimethyl fumarate (HR=0.78, 95% CI=0.66 to 0.92). Further, natalizumab (HR=0.43, 95% CI=0.32 to 0.56) showed a superior ATE in reducing disability worsening and in disability improvement (HR=1.32, 95% CI=1.08 to 1.60). The pairwise ATT comparisons also showed superior effects of natalizumab followed by fingolimod on relapses and disability.

WHAT IS ALREADY KNOWN ON THIS TOPIC
⇒ Direct simultaneous comparisons of several multiple sclerosis (MS) therapies, outside of analyses of secondary data—such as network meta-analyses, are presently lacking.

WHAT THIS STUDY ADDS
⇒ We applied methods of causal inference to emulate a multiarm randomised trial comparing simultaneously the effectiveness of six commonly used disease-modifying therapies and no treatment among patients with relapsing-remitting MS.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY
⇒ This study fills the gap in the lack of methodology for robust simultaneous comparison of multiple disease-modifying therapies in MS, especially where randomised trials for such comparisons are not feasible.

Conclusions The effectiveness of natalizumab and fingolimod in active RRMS is superior to dimethyl fumarate, teriflunomide, glatiramer acetate and interferon beta. This study demonstrates the utility of MSM in emulating trials to compare clinical effectiveness among multiple interventions simultaneously.
INTRODUCTION

Disease-modifying therapies (DMTs) play a crucial role in the management of multiple sclerosis (MS). DMTs reduce risks of relapses, disability worsening, secondary progression and delay mortality. 1-4 Multiple therapies have been approved for the treatment of MS, 5 and neurologists and patients can now choose from among different mechanisms of action and methods of administration. Therefore, evidence comparing effectiveness to guide the choice among DMTs is needed.

Magnitudes of the effects of different DMTs when compared with placebo, as reported by meta-analyses of randomised controlled trials, suggest that there are differences in the efficacy of these DMTs. 6 Randomised clinical trials involving multiple DMTs are restricted by significant financial and practical limitations. Because of methodological challenges, comparisons of treatments in MS cohorts and registries have been limited to pairwise and triple comparisons. The difficulty in matching more than three groups leads to a dramatic loss of power and hence limited generalisability. Simultaneous comparisons of multiple DMTs in a single trial design are still lacking.

To overcome these limitations, we aimed to implement an extension of marginal structural models (MSMs) to simultaneously compare the effectiveness of multiple MS treatments across a broad spectrum of patients with MS. To this end, we have extended a methodology developed in cross-sectional data to longitudinal data with time-varying covariates. 7

MATERIALS AND METHODS

Ethics statement

Data were obtained from the international MSBase registry (registered with WHO ACTRN12605000455662). 8 Participants were required to provide informed consent as per the local regulations.

Study design

This study emulates a randomised trial comparing relapse frequency, disability worsening and disability improvement among six different DMTs, using data from MSBase (online supplemental table 1). Patients were followed from their first recorded MS clinic visit with disability data (study baseline). For each patient, the first recorded treatment epoch was analysed. Patients were censored at the time of switching/discontinuing DMT or the last recorded follow-up (online supplemental figure 1). If a patient was untreated during the first recorded epoch, this epoch was included in the analysis as ‘untreated’. The primary analysis estimated the average treatment effect (ATE)—a hypothetical effect of treatment if the entire study population had been treated with a DMT of interest versus glatiramer acetate. 9,10 Because the reference pseudo-cohort in ATE is constant, it enabled us to combine multiple pairwise comparisons of the index and reference DMTs to compare the effectiveness across multiple index DMTs simultaneously. Its estimation of the treatment effect across the entire study sample offers maximal generalisability of the result. The secondary analysis estimated the average treatment effect among the treated (ATT)—a comparison between a factual (observed) effect of a DMT among patients who were treated with that DMT and a counterfactual (not observed) effect of another DMT in the same study subpopulation. 7 Therefore, the results of ATT provide information about the pairwise comparative effectiveness of different DMTs in clinical contexts in which they are typically used.

Study population

In March 2019, we extracted data from 23,236 patients from 74 centres across 35 countries. The follow-up was recorded prospectively since 1 January 2006 as part of clinical practice following MSBase Study Protocol (available from www.msfbase.org). We excluded centres in the lowest quintiles of data quality or generalisability. 11 Included patients who were diagnosed with clinically isolated syndrome or relapsing-remitting MS 12,13 and were censored at the conversion to secondary progressive MS as per their neurologists. 14 The minimum required data consisted of recorded follow-up ≥1 year, ≥3 Expanded Disability Status Scale (EDSS) scores with ≥1 score recorded per year, sex, date of birth, MS onset date and MS course. We conducted a sensitivity analysis restricted to patients with available MRI information.

Variables of interest

Follow-up time was divided into intervals (bins) of 6 months. Bins with DMT recorded for ≥15 days were considered as treated. Potential confounders of treatment effect were captured in each follow-up bin (see the Statistical methods section).

Disability was quantified with the EDSS, excluding scores obtained ≤30 days after a relapse. Neurostatus EDSS certification was required at the participating centres. 11 Where no EDSS scores were recorded within a bin, the preceding EDSS score was carried over (only 27% of bins). In our previous work in this cohort, EDSS scores in consecutive bins were highly correlated (r=0.95). 16

The presence/absence of relapses, confirmed disability worsening and disability improvement during each bin, were binary variables. Relapses were recorded by treating neurologists as new symptoms or exacerbation of existing symptoms persisting for ≥24 hours, in the absence of concurrent illness/fever, and occurring ≥30 days after a previous relapse. Disability worsening was defined as an increase in EDSS by 1 step (1.5 steps if baseline EDSS=0 and 0.5 steps if EDSS >5.5) confirmed by subsequent EDSS scores over ≥12 months, as over 80% of such events correspond to long-term worsening of disability. 17 Disability improvement was defined as a decrease in EDSS by 1 step (≤1.5 steps if baseline EDSS ≤1.5 and 0.5 steps if EDSS >6) confirmed over ≥12 months. No carried-over EDSS scores were used for evaluation of disability worsening or improvement. The presence/absence of new or enlarging T2 hyperintense lesions or contrast-enhancing lesions on cerebral MRI was reported by treating neurologists in some patients.

Data availability

Data Dictionary and MSBase Study Protocol are available from www.msfbase.org. MSBase is a data processor. Data access to external parties can be granted at the discretion of each MSBase Principal Investigator (data controller), who will need to be approached individually for permission.

Statistical methods

We used generalised boosted regression models (GBMs) to estimate inverse probability of treatment weight (IPW) and inverse probability of censoring weight (IPCW). Machine learning methods have been shown to outperform simple regression methods for iterative variable selection. 18-20 GBMs can adjust for many covariates and capture complex and non-linear relationships without overfitting.

To avoid loss of information that would limit the generalisability of our results, all treatments including those with a small


1005
Multiple sclerosis

number of users (such as alemtuzumab, mitoxantrone and ocrelizumab) were included when calculating the ATE weights. These DMTs with a small number of patients were, thereafter, excluded before the analysis of head-to-head comparisons of therapies. Data were analysed with R version 3.6.3, using the package GBM.

Estimation of ATE weights

ATE comparing treatment A and treatment B is estimated as:

\[ \text{ATE} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{m_i} \sum_{t=1}^{m_i} \left( Y_{it}^A - Y_{it}^B \right) \right) \]

where \( Y_{it}^A \) and \( Y_{it}^B \) represent time-varying confounding variables and \( t \) is time. The probability for a patient with characteristics \( X \) of receiving treatment \( t \) was computed for each dummy treatment indicator at each 6-month bin given their covariate history (age, sex, pregnancy status, treatment history, history of relapses, MS duration, EDSS scores, MS course and year of birth). Stabilised weights for each patient \( i \) for each given treatment \( t \) at bin \( k \) were calculated as:

\[ w_{ik}[t] = \prod_{k=0}^{t} \frac{P(A_k = a_k | [A_{k-1} = a_{k-1}], B = b_k)}{P(A_k = a_k | [A_{k-1} = a_{k-1}], B = b_k, X_{kp} = x_{kp})} \]

where \( A \) is the treatment status at bin \( k \), \( B \) represents fixed confounding variables and \( L \) represents time-varying confounding variables. Thus, the weights reflect the probability of patients’ treatment status at each bin given their demographic and disease history.

To account for censoring due to treatment switch, conversion to secondary progressive MS or study dropout, we calculated stabilised IPCWs as a function of baseline and time-varying covariates.

\[ c_{ik}[t] = \prod_{k=0}^{t} \frac{P(A_k = a_k | [A_{k-1} = a_{k-1}], B = b_k)}{P(A_k = a_k | [A_{k-1} = a_{k-1}], B = b_k, X_{kp} = x_{kp})} \]

Estimation of ATT weights

Let \( \mu^t \), \( \mu^t \) equal the counterfactual means for patients who received treatment \( t \) had they instead received treatment \( t' \), \( \mu^t \), \( \mu^t \), where the ATT of treatment \( t \) relative to \( t' \) is the difference between the factual and the counterfactual means, \( \mu^t - \mu^t \).

We used a subsample of the population with \( t \) and one alternative treatment \( t' \) to calculate the weight for a patient receiving the target treatment \( t \) as the ratio of the probability of receiving the treatment actually received (\( t' \)) and the probability of receiving \( t' \). The procedure is repeated for each treatment \( t \). Each patient is weighted by the inverse probability of receiving the treatment they actually received (\( t' \)) relative to the probability of receiving the alternative treatment (\( t' \)). Like ATE weights ATT weights are the product of the IPWs and IPCWs to take into account censoring due to treatment switch or study dropout.

Comparisons of treatment effects

Baseline characteristics of the study population were described using mean (SD) for continuous variables and n (%) for categorical variables. The standardised bias or absolute standardised mean difference was used as a balance metric to assess covariate balance between treatment groups before and after weighing (overall and per bin). For a covariate \( k \) and a treatment \( t \), the balance metric is computed as:

\[ PSB_{ik} = |\bar{X}_{ik} - \bar{X}_{kp}|/\text{SD}_{ik} \]

where \( PSB_{ik} \) is the population standardised bias, \( X_{ik} \) represents the mean of the covariate \( k \) for the treatment \( t \), \( X_{kp} \) is the mean of the covariate in the population and \( \text{SD}_{ik} \) is population SD for the covariate \( k \).

We used Cox proportional hazards models with the IPWs (MSM Cox model) to compare hazards of relapses, disability worsening and disability improvement among pseudo-cohorts treated with sufficiently represented therapies versus the reference therapy. For the ATE and ATT, the main reference was ‘glatiramer acetate’, with sensitivity analyses conducted against alternative references—fingolimod and natalizumab. In addition, the weighted models were further adjusted for MS duration, EDSS or age when the covariate balance between weighted groups exceeded 0.20. Sufficiently represented therapies were defined as treatments administered to at least 200 eligible individuals.

RESULTS

Among 23,236 eligible patients in sufficiently represented treatment groups, the mean follow-up was 2.8 years (SD 2.2) with 6.5 EDSS visits (SD 6.3). At baseline (first recorded EDSS visit), interferon beta was the most common treatment (41%), followed by glatiramer acetate (11%, table 1). Fingolimod and natalizumab represented 4.5% and 5%, respectively, while dimethyl fumarate (1.5%) and teriflunomide (1.3%) were less common. Thirty-five per cent of eligible patients were untreated at baseline. Patients on teriflunomide tended to be older than those untreated, on interferon beta or on natalizumab. MS duration at baseline tended to be marginally longer for fingolimod, natalizumab, glatiramer acetate and teriflunomide and shorter in untreated patients. Baseline disability was greater in those treated with natalizumab than in the other treatment categories. Patients on natalizumab and fingolimod had used more DMTs before becoming eligible for this study and were more likely to have experienced relapses on DMTr in the pre-baseline year. More than 900 patients were still treated with their index study therapy at 5 years from its commencement, allowing sufficient power for a comparison of treatment effectiveness over the 5-year period (online supplementary figure 2). When compared with participants included in this study, those who were excluded tended to be older, with longer MS duration at baseline, higher levels of disability, fewer MRI lesions and were more likely untreated (online supplemental table 2).

Covariate balance

Weights from the ATE models significantly improved covariate balance between each treatment and the pooled sample of alternative treatment states at individual 6-monthly bins (online supplemental figure 3) as well as overall (figure 1). Sufficient balance was maintained throughout most of the follow-up while the treatment groups were sufficiently populated. Standardised differences of <0.20 are considered acceptable, resulting in ≥92% overlap between compared cohorts. The covariates that were challenging to balance and showed standardised mean differences marginally >0.20 were EDSS for the natalizumab versus other therapies and relapses in the prior 12 months for fingolimod versus other therapies (online supplemental figures 3 and 4). The models were adjusted for these residual imbalances.


Downloaded from http://jnnp.bmj.com/ on 6 July 2023 by guest. Protected by copyright.
The greater effect on reducing relapses than the other therapies—glatiramer acetate, fingolimod and natalizumab (HR=0.60, 95% CI=0.54 to 0.66), dimethyl fumarate (HR=1.15, 95% CI=0.96 to 1.35) and no-treatment (HR=0.89, 95% CI=0.82 to 1.05) (figure 5). The three sets of comparisons using different treatment were first weighted to match those of patients included in the weights. Results were similar to those observed in the main analysis. However, the superiority of natalizumab had a greater effect on reducing relapses than the other therapies. The relative effects of dimethyl fumarate (−22%; HR=0.78, 95% CI=0.66 to 0.92), teriflunomide (−11%; HR=0.89, 95% CI=0.75 to 1.06) and interferon beta (−5%; HR=0.95, 95% CI=0.89 to 1.00) were quantitatively smaller and no-treatment was detrimental for relapse incidence (+33%; HR=1.35, 95% CI=1.27 to 1.44). Natalizumab was associated with a lower cumulative hazard of confirmed disability worsening than the other therapies. The hazards of disability worsening relative to glatiramer acetate were: natalizumab (HR=0.43, 95% CI=0.32 to 0.56), fingolimod (HR=0.85, 95% CI=0.67 to 1.06), dimethyl fumarate (HR=0.86, 95% CI=0.51 to 1.47), teriflunomide (HR=0.56, 95% CI=0.31 to 0.99), interferon beta (HR=1.08, 95% CI=0.96 to 1.23) and no-treatment (HR=1.04, 95% CI=0.89 to 1.21). Relative to glatiramer acetate, natalizumab (HR=1.32, 95% CI=1.08 to 1.60) led to a higher probability of disability improvement compared with fingolimod (HR=1.18, 95% CI=0.96 to 1.46), dimethyl fumarate (HR=1.15, 95% CI=0.82 to 1.60), teriflunomide (HR=0.70, 95% CI=0.44 to 1.11), interferon beta (HR=1.03, 95% CI=0.91 to 1.18) and no treatment (HR=0.91, 95% CI=0.78 to 1.05) (figure 2).

We conducted sensitivity analyses that were restricted to patients with MRI data, where the number of MRI lesions was included in the weights. Results were similar to those observed in the main analysis. However, the superiority of natalizumab for disability improvement was attenuated (online supplementary figure 5). The three sets of comparisons using different reference therapies—glatiramer acetate, fingolimod and natalizumab, showed results consistent with the primary analysis.

**Pairwise comparisons of therapies: ATT models**

In the ATT approach, characteristics of patients using other treatments were first weighted to match those of patients

---

**Table 1 Unweighted characteristics of study participants at the first visit**

<table>
<thead>
<tr>
<th></th>
<th>None (n=1816)</th>
<th>Glatiramer acetate (n=2631)</th>
<th>Interferon beta (n=1039)</th>
<th>Fingolimod (n=5956)</th>
<th>Natalizumab (n=1168)</th>
<th>Dimethyl fumarate (n=3533)</th>
<th>Teriflunomide (n=312)</th>
<th>Overall (n=23236)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age, mean (SD)</strong></td>
<td>35.5 (10.5)</td>
<td>38.2 (10.2)</td>
<td>36.2 (10.4)</td>
<td>37.1 (10.4)</td>
<td>36.0 (10.5)</td>
<td>36.9 (11.7)</td>
<td>42.0 (11.6)</td>
<td>36.3 (10.5)</td>
</tr>
<tr>
<td><strong>Female sex, n (%)</strong></td>
<td>5907 (72.3)</td>
<td>1919 (72.9)</td>
<td>6846 (71.6)</td>
<td>748 (72.0)</td>
<td>813 (69.6)</td>
<td>225 (63.7)</td>
<td>227 (72.8)</td>
<td>16685 (71.8)</td>
</tr>
<tr>
<td><strong>MS duration, mean (SD)</strong></td>
<td>4.96 (6.56)</td>
<td>7.03 (6.97)</td>
<td>6.72 (6.74)</td>
<td>7.24 (7.07)</td>
<td>7.49 (6.97)</td>
<td>6.10 (7.27)</td>
<td>8.23 (7.90)</td>
<td>6.21 (6.82)</td>
</tr>
<tr>
<td><strong>EDSS, mean (SD)</strong></td>
<td>2.03 (1.48)</td>
<td>2.18 (1.52)</td>
<td>2.06 (1.44)</td>
<td>2.22 (1.56)</td>
<td>3.02 (1.68)</td>
<td>1.87 (1.43)</td>
<td>2.02 (1.50)</td>
<td>2.12 (1.50)</td>
</tr>
<tr>
<td><strong>RRMS, n (%)</strong></td>
<td>6206 (76.0)</td>
<td>2459 (93.5)</td>
<td>8601 (89.9)</td>
<td>1015 (97.7)</td>
<td>1154 (98.8)</td>
<td>337 (95.5)</td>
<td>296 (94.9)</td>
<td>20 068 (86.4)</td>
</tr>
<tr>
<td><strong>Relapses (last 3 months), n (%)</strong></td>
<td>5675 (82.2)</td>
<td>2078 (79.0)</td>
<td>7627 (79.7)</td>
<td>832 (80.1)</td>
<td>874 (74.8)</td>
<td>269 (76.2)</td>
<td>256 (82.1)</td>
<td>18 651 (80.3)</td>
</tr>
<tr>
<td><strong>Number of previous DMTs, n (%)</strong></td>
<td>7155 (87.6)</td>
<td>2138 (81.3)</td>
<td>8254 (86.3)</td>
<td>664 (63.9)</td>
<td>537 (46.0)</td>
<td>257 (72.8)</td>
<td>225 (72.1)</td>
<td>19 230 (82.8)</td>
</tr>
<tr>
<td><strong>Number of MRI lesions (last 12 months), n (%)</strong></td>
<td>623 (76.0)</td>
<td>355 (13.5)</td>
<td>968 (10.1)</td>
<td>210 (20.6)</td>
<td>67 (5.9)</td>
<td>22 (7.6)</td>
<td>162 (5.6)</td>
<td>236 (11.2)</td>
</tr>
<tr>
<td><strong>Relapses on DMT (during last 12 months), n (%)</strong></td>
<td>390 (4.8)</td>
<td>138 (5.2)</td>
<td>343 (3.6)</td>
<td>161 (15.5)</td>
<td>349 (29.5)</td>
<td>33 (9.3)</td>
<td>22 (7.1)</td>
<td>1431 (6.2)</td>
</tr>
<tr>
<td><strong>Relapses on high-efficacy DMTs (during last 12 months), n (%)</strong></td>
<td>8074 (98.8)</td>
<td>2331 (88.6)</td>
<td>8332 (87.1)</td>
<td>911 (87.7)</td>
<td>954 (81.7)</td>
<td>336 (95.2)</td>
<td>288 (92.3)</td>
<td>21 226 (91.3)</td>
</tr>
<tr>
<td><strong>Severe relapses (last 12 months), n (%)</strong></td>
<td>79 (1.0)</td>
<td>237 (9.0)</td>
<td>1007 (10.5)</td>
<td>110 (10.6)</td>
<td>155 (13.3)</td>
<td>15 (4.2)</td>
<td>22 (7.1)</td>
<td>1625 (7.0)</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td>2332 (99.9)</td>
<td>8662 (98.6)</td>
<td>9624 (100.0)</td>
<td>1008 (97.0)</td>
<td>1137 (97.3)</td>
<td>352 (99.7)</td>
<td>309 (99.0)</td>
<td>23 162 (99.7)</td>
</tr>
<tr>
<td><strong>DMDs, disease-modifying therapies; EDSS, Expanded Disability Status Scale; MS, multiple sclerosis; RRMS, relapsing-remitting multiple sclerosis.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Multiple sclerosis

using glatiramer acetate (reference therapy). ATT results are pairwise comparisons between therapies generalisable only to the subpopulation treated with the therapy of interest (active therapy). Results suggest that patients on glatiramer acetate would have experienced fewer relapses if they were instead treated with several other treatments (figure 3): natalizumab (−32%; HR=0.68, 95% CI=0.60 to 0.78), fingolimod (−38%; HR=0.62, 95% CI=0.55 to 0.70) or interferon beta (−7%; HR=0.93, 95% CI=0.87 to 0.98). Patients treated with glatiramer acetate would also have experienced fewer events of disability worsening if they were treated with natalizumab (−39%; HR=0.61, 95% CI=0.44 to 0.86) and a greater chance of disability improvement if they were treated with natalizumab (HR=1.64, 95% CI=1.32 to 2.05) or fingolimod (HR=1.29, 95% CI=1.02 to 1.63).

Second, the characteristics of patients on other treatments were weighted to match those who were treated with natalizumab (reference category). Results suggest that patients on natalizumab would have experienced higher risks of relapses if they had been treated with any other of the included therapies. They would have had a lower chance of disability improvement if they were treated with the other therapies except for dimethyl fumarate, and a higher risk of disability worsening if they were treated with interferon beta (figure 3).

Third, fingolimod was set as a reference category and characteristics of patients treated with other treatments were weighted to match those of patients on fingolimod. Results indicate that patients who were treated with fingolimod would have experienced more relapses on other treatments, except for natalizumab (for which there was no evidence of a difference in this patient group). The chances for disability improvement in patients treated with fingolimod would have been higher if they were treated with natalizumab (HR=1.39, 95% CI=1.07 to 1.80) and lower if they were treated with teriflunomide, glatiramer acetate or if they were untreated. The only therapy for which we have found an indication of a difference in the hazard of disability worsening compared to fingolimod is dimethyl fumarate (HR=0.69, 95% CI=0.50 to 0.95).

Figure 2  Comparison of multiple therapies using ATE. Results from Cox proportional hazards models with the inverse probability of treatment weights to compare hazards of relapses, disability worsening and disability improvement among pseudo-cohorts treated with sufficiently represented therapies versus the reference therapy. ATE, average treatment effect.
worsening was teriflunomide, in which the risk of worsening would have been lower if it was used in the group treated with fingolimod.

**DISCUSSION**

We have emulated a quasi-randomised trial simultaneously comparing effectiveness among six commonly used DMTs for MS and untreated state. Among the selected therapies compared with the ATE modelling approach, natalizumab ahead of fingolimod is most effective in preventing relapses, disability worsening and disability improvement for pairwise comparisons of treatment groups. ATT, average treatment effects among the treated.

The pairwise comparisons corroborated the overall observations, supporting the superiority of natalizumab and, in some instances, fingolimod over the other included therapies in controlling MS activity. More specifically, among patients treated with natalizumab, who, together with patients treated with fingolimod, tended to have experienced more severe disease, we identified not only a superior effect of natalizumab on relapses but also a tendency to reduce the frequency of disability worsening and disability improvement for pairwise comparisons of treatment groups. ATT, average treatment effects among the treated.

The available results from randomised controlled trials mostly evaluate the efficacy of therapies compared with placebo over 2–3 years. Our present study not only corroborates the results of trials but also extends the conclusions to multiple direct pairwise comparisons and multiarm comparisons of the studied interventions over 5 years. While one meta-analysis of randomised trials did not report a superiority of natalizumab over fingolimod in reducing relapses over 2 years, another meta-analysis found evidence for a superiority of natalizumab over fingolimod for relapse outcomes and disability improvement. In keeping with our findings, the risks of relapse in both above-cited meta-analyses were lower in natalizumab or fingolimod compared with interferon beta, glatiramer acetate, dimethyl fumarate and teriflunomide. Pairwise comparisons of relapse frequencies between glatiramer acetate, interferon beta and teriflunomide did not find evidence for differences. The risk of 3-month confirmed disability worsening tended to be lower in natalizumab compared with any other DMTs included in the meta-analyses, reaching the pre-defined threshold for statistical significance for glatiramer acetate and interferon beta-1b.

In addition to the data from randomised trials and their meta-analyses, data from MS registries and cohorts have been used for pairwise or three-way comparisons between groups treated with some of the DMTs, balanced with propensity score matching or weighting. A study in the registry of the Swiss Federation for Common Tasks of Health Insurances reported the comparative effectiveness of natalizumab versus fingolimod over a median follow-up of 1.8 years in patients who switched therapy from interferon beta or glatiramer acetate. Patients who switched DMT to natalizumab had lower risks of relapses and were more likely to experience disability improvement while no differences were found in the proportion of patients free from EDSS progression. We have previously reported similar results for natalizumab versus fingolimod using the MSBase data over a mean 12-month follow-up. These conclusions were later corroborated in a large analysis combining data from three large registries—OFSEP, MSBase and Danish MS Registry. Several other observational studies reported similar relapse incidence among patients treated with fingolimod, dimethyl fumarate or teriflunomide, but with a small relative advantage among patients switching from glatiramer acetate or interferon beta to fingolimod versus dimethyl fumarate or teriflunomide (absolute difference of approximately 0.1 relapse per year).

The application of the novel methodological approach, combining marginal structural models for estimation of causal associations between time-varying exposures to multiple therapies with generalised boosted models builds on the use of different components of this methodology in previous studies. Similar to the individual components of this design, the current design is broadly applicable to data from various settings, ranging from cohorts, registries and databases to repurposed data from randomised controlled trials.

The main limitation of this study is the risk of potential unmeasured confounding due to missing data—especially MRI information. Reassuringly, sensitivity analyses within a subgroup with MRI data available support the primary ATE analysis. While this study allows generalisation of its observations to a population of patients with a relatively active MS, the restricted size of the subgroups with low disease activity or diagnosed with clinically isolated syndrome did not allow us to explore the effectiveness of therapies specifically in these clinical scenarios. A potential heterogeneity in the acquisition of relapses and disability...
information represents another limitation. We have mitigated this risk by prospectively defining relapse in the MSBase Study Protocol and by requiring Neurostatus certification at each site. Further, our models cannot account for the delayed effects of previous treatments in preventing relapses and disability. To mitigate this limitation, we have used cumulative weights that use information about treatment exposure during previous bins included in the analysis. Potential effects of regression to the mean were equalised across the compared treatments by periodically adjusting the cumulative IPW and IPCW for the number of relapses recorded during the preceding 6 months. Informed censoring leading to attrition bias may represent a limitation where retention of longitudinal data in the study is conditional on persistence of study therapy. This was accounted for by weighting on the inverse probability of censoring. It is encouraging that our weighting strategy resulted in good covariate balance among treatments suggesting that MSMs are suitable for simultaneous comparisons of several MS therapies in the context of their complex causal relationships.

CONCLUSIONS
This study for the first time simultaneously compares and quantifies the effectiveness of six commonly used MS therapies and no treatment. Its results suggest that among patients with active relapsing MS, natalizumab followed by fingolimod offers an advantage over dimethyl fumarate, teriflunomide, interferon β and glatiramer acetate in reducing the risk of relapses and improving disability outcomes. However, the individual use of these therapies should be also guided by specific clinical scenarios and their safety profiles, which were not evaluated in this study. The methodological development proposed in this report allows simultaneous comparison of the effectiveness of multiple treatments using observational data while accounting for time-dependent confounding and attrition bias. We highlight the conceptual differences between the ATE and ATT approaches to causal inference and demonstrate how complementary clinical questions are answered by utilising both strategies. This approach allows us to inform neurologists about both the overall relative effectiveness of the compared therapies as well as comparisons between selected treatment pairs in populations that are typically treated with the therapy of interest.

Author affiliations
1CORe, Department of Medicine, The University of Melbourne, Melbourne, Victoria, Australia
2Health and Biosecurity Unit, Commonwealth Scientific and Industrial Research Organisation, Melbourne, Victoria, Australia
3Neuroimmunology Centre, Department of Neurology, The Royal Melbourne Hospital City Campus, Parkville, Victoria, Australia
4Department of Neurology, Center of Clinical Neuroscience, Charles University, Prague, Czech Republic
5General University Hospital in Prague, Prague, Czech Republic
6Department of Medical and Surgical Sciences and Advanced Technologies, University of Catania ‘G.F. Ingrassia’, Catania, Italy
7Department of Neurology, Isfahan University of Medical Sciences, Isfahan, Iran (the Islamic Republic of)
8Dokuz Eylul University, Izmir, Turkey
9Neurology, Hospital Universitario Virgen Macarena, Sevilla, Spain
10Department of Neuroscience, Imaging, and Clinical Sciences, Gabriele d’Annunzio University of Chieti and Pescara, Chieti, Italy
11UO5I Riabilitazione Scelosi Multipla, IRCCS Istituto delle Scienze Neurologiche di Bologna, Bologna, Italy
12Dipartimento di Scienze Biomediche e Neuromotorie, Università di Bologna, Bologna, Italy
13Division of Neurology, Department of Medicine, Amiri Hospital, Kuwait City, Kuwait
14CHUM MS Center, Montreal, Quebec, Canada
15Universite de Montreal, Montreal, Quebec, Canada
16School of Medicine, Ondokuz Mayis Universities, Samsun, Turkey
17KTU Medical Faculty Farabi Hospital, Trabzon, Turkey
18Neuro Rive-Sud, Greenfield Park, Quebec, Canada
19Department of Neuroscience, Azienda Ospedaliero-Università di Modena, Modena, Italy
20CISS de Chaudriére-Appalaches, Levis, Quebec, Canada
21Nehme and Therese Tohme Multiple Sclerosis Center, American University of Beirut Medical Center, Beirut, Lebanon
22Department of Neurology, American University of Beirut, Beirut, Lebanon
23Department of Neurology, Koc Universities, Istanbul, Turkey
24Koc University Research Center for Translational Medicine, Istanbul, Turkey
25Department of Neurology, Zuyderland Medical Centre Sittard-Geleen, Sittard-Geleen, The Netherlands
26School for Mental Health and Neuroscience, Maastrict University, Maastricht, The Netherlands
27University of Newcastle Hunter Medical Research Institute, Newcastle, New South Wales, Australia
28Department of Neurology, John Hunter Hospital, Newcastle, New South Wales, Australia
29Foundation National Neurological Institute C Mondino Institute for Hospitalization and Care Scientific, Pavia, Italy
30Department of Neurology, Faculty of Medicine, Hacettepe University, Ankara, Turkey
31Ospedali Riuniti di Salerno, Salerno, Italy
32Department of Neurology, St Vincent’s University Hospital, Dublin, Ireland
33UOC Neurologia, Azienda Sanitaria Unica Regionale Marche - AV3, Macerata, Italy
34Royal Victoria Hospital, Belfast, UK
35Department of Neurology, Centro Hospitalar de Sáo João, Porto, Portugal
36Department of Neurology, ASL3 Genovese, Genova, Italy
37Department of Rehabilitation, Casa di Cura Centro di Recupero e Rieducazione Funzionale Mons Luigi Novarese, Monciglieto, Italy
38Research Center for Clinical Neuroimmunology and Neurosciences Basel (RC2NB) and MS Center, Neurologic Clinic and Polyclinic, Departments of Head, Spine and Neuroimmunology and Clinical Research, University Hospital Basel, Basel, Switzerland
39Department of Basel, Basel, Switzerland
40Liverpool Hospital, Sydney, New South Wales, Australia
41Flinders University, Adelaide, South Australia, Australia
42Department of Medicine and Surgery, University of Parma, Parma, Italy
43Groene Hart ziekenhuis, Gouda, The Netherlands
44UQCCR, The University of Queensland, Brisbane, Queensland, Australia
45Neurocenter Jihlava, Jihlava, Czech Republic
46Department of Neurology, Alfred Hospital, Melbourne, Victoria, Australia
47Department of Neurology, Neurosciences, School of Medicine, The University of Melbourne, Melbourne, Victoria, Australia
48Department of Neurology, Galdakao-Uxamsolo Hospital, Galdakao, Spain
49Westmead Hospital, Westmead, New South Wales, Australia
50Universitair Ziekenhuis Gent, Gent, Belgium
51Department of Neurology, Razi University Hospital, Tusn, Tunisia
52Department of Neurology, University of Tunis El Manar, Tunis, Tunisia
53Instituto de Investigacion Sanitaria Biodonostia, Hospital Universitario de Donostia, San Sebastian, Spain
54South and East Belfast Health and Social Services Trust, Belfast, UK
55Neurology, Hospital Universitario Reina Sofia, Cordoba, Spain
56Department of Medicine, Sultan Qaboos University Hospital, Seeb, Oman
57Sultan Qaboos University, Muscat, Oman
58University Hospital Geelong, Geelong, Victoria, Australia
59Department of Neurology, King Fahad Specialist Hospital-Dammam, Khoor, Saudi Arabia
60Department of Neurology, University of Debrecen, Debrecen, Hungary
61Department of Neurology, Hospital General de Alicante, Alicante, Spain
62Department of Neurology, Instituto Nacional de Ciencias Medicas y Nutricion Salvador Zubiran, Mexico City, Mexico
63Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland, USA

Twitter Alessandra Lugaresi @ALugaresi and Eduardo Aguera-Morales @doctoredu

Contributors ID performed the analyses, interpreted the results, drafted and reviewed the manuscript. TK obtained funding, conceptualised the study, contributed data, designed the analyses, interpreted the results, edited and reviewed the manuscript and acts as the study guarantor. CBM, SS, IR, EAS contributed to the study design, interpreted the results and edited and reviewed the manuscript. DH, EKHP, FP, VS, SQ, SE, MO, AL, RAI, AP, PD, MT, CB, FGM, PD, DS, PG, BY, AA, OG, JLS, RB, RK, Glu, CM, EC, SHU, NIS, CSO, LK, SHO, MS, FG, KiG, PAM, RAM, AvwB, HW, JLS, M, V, GlA, YS, RS, GC, TC, TO, EG, EAM, AA-A, CSH, TMA-H, TC, APS, IF, contributed data, interpreted the results and edited and reviewed the manuscript.

Funding This study was financially supported by National Health and Medical Research Council of Australia [1129189, 1140766, 1080518].
Competing interests  The authors report the following relationships: speaker honoraria, advisory board or steering committee fees, research support and/or conference travel support from Actelion (EKH, RA), Almirall (MT, FG, RB, CRT, JLS-M), Bayer (MT, AL, PS, RA, MT, CB, JS, EP, VPR, RB, DS, RA, JO, JLSM, SH, CR, AGK, TC, NS, BT, MS, CAS), Biotest (TL, AGK, BT), Biogen (TK, DS, DH, EK, MT, GI, AL, MG, PD, PG, VI, AVN, FG, PS, DF, RA, RH, CB, JLS, ER, VPR, FG, RB, RA, CRT, JP, JO, MB, JLSM, SH, CR, CS, OGerlach, AGK, TC, BS, NS, BT, MS, HB), Biologix (RA), BMS/Celgene (EKH, AL), Genpharm (RA), Genzyme-Sanoﬁ (TK, EK, MT, GI, AL, MG, PD, AVW, FG, PS, DF, RA, RH, MT, CB, JLS, EP, Ep, EP, VPR, FG, RB, DS, CRT, JP, JO, MB, JLSM, SH, O Gerlach, AGK, HB), GSK (RA), Innate Immunotherapeutics (AGK), Lundbeck (EP), Merck / EMD (TK, DH, EK, MT, GI, AL(Merck Serono), MG, PD, PG, VI, AVW, PS, DF, RA, RH, MT, CB, JLS, EP, VPP, FG, RB, DS, RA, JO, MB, JLSM, CR, FM, O Gerlach, AGK, TC, BS, MS, HB), Mitsubishi (FG), Novartis (TK, TS, DH, EK, MT, GI, AL, MG, PD, PG, VI, AVN, FG, PS, DF, RA, RH, MT, CB, JLS, EP, VPP, FG, RB, DS, RA, CRT, JP, JO, MB, JLSM, SH, CR, FM, CS, OGH, AGK, TC, BS, NT, MS, HB), ONO Pharmaceuticals (FG), Roche (TK, EK, AL, MT, CB, VPP, BT), Teva (TK, DH, EK, MT, GI, AL, MG, PD, PG, VI, FG, PS, DF, RH, MT, CB, JLS, VPP, RB, RS, RA, JP, JO, JLSM, CR, AGK, TC, MS, CAS), WebMD (TK), UCB (EP).

Patient consent for publication  Not applicable.

Ethics approval  This study involves human participants and was approved by Melbourne Health (2006.044). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review  Not commissioned; externally peer reviewed.

Data availability statement  Data are available upon reasonable request. Not commissioned; externally peer reviewed.

Patient consent for publication  Not applicable.

ORCID iDs

Charles B Malpas http://orcid.org/0000-0003-0524-3718
Sifat Sharmin http://orcid.org/0000-0001-7294-9723
Izanne Roos http://orcid.org/0000-0003-0371-3666
Sara Eichau http://orcid.org/0000-0001-9159-3128
Marco Onofri http://orcid.org/0000-0002-0480-2495
Alessandra Lugaresi http://orcid.org/0000-0003-2902-5589
Diana Ferraro http://orcid.org/0000-0002-4852-5207
Claudio Solaro http://orcid.org/0000-0002-6713-4623
Ludwig Kappos http://orcid.org/0000-0003-4175-5509
Pamela A McCombe http://orcid.org/0000-0003-2704-8517
Anneke van der Walt http://orcid.org/0000-0001-9615-3797
Tamara Castillo-Trivino http://orcid.org/0000-0002-9249-3185
Tomas Kalinick http://orcid.org/0000-0003-3778-1376

REFERENCES

13 Polman CH, Wolinsky JS, Reingold SC. Multiple sclerosis diagnostic criteria: three years later. Mult Scler 2011;17:11–12.
## Online supplement

### Supplementary Table 1. Comparison of patients eligible for and excluded from the present study.

<table>
<thead>
<tr>
<th></th>
<th>Excluded (n=38574)</th>
<th>Included (n=23236)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age first visit, Mean (SD)</strong></td>
<td>39.1 (12.6)</td>
<td>36.3 (10.5)</td>
</tr>
<tr>
<td><strong>Female sex</strong></td>
<td>26899 (69.7%)</td>
<td>16685 (71.8%)</td>
</tr>
<tr>
<td><strong>MS duration at baseline, Median [Q1, Q3]</strong></td>
<td>7.53 (8.59)</td>
<td>6.21 (6.82)</td>
</tr>
<tr>
<td><strong>Disability</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDSS 0-3.5</td>
<td>22094 (57.3%)</td>
<td>20236 (87.1%)</td>
</tr>
<tr>
<td>EDSS 4-5.5</td>
<td>5086 (13.2%)</td>
<td>2230 (9.6%)</td>
</tr>
<tr>
<td>EDSS 6-9.5</td>
<td>6087 (15.8%)</td>
<td>770 (3.3%)</td>
</tr>
<tr>
<td>Missing</td>
<td>5307 (13.8%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Presence of new/enlarging/enhancing cerebral MRI lesions during the last 12 months</td>
<td>913 (2.4%)</td>
<td>1314 (5.7%)</td>
</tr>
<tr>
<td>Receiving disease modifying therapy at baseline</td>
<td>9736 (25.2%)</td>
<td>15068 (64.8%)</td>
</tr>
</tbody>
</table>
**Supplementary table 2**: Summary of the study protocol

<table>
<thead>
<tr>
<th>Protocol component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>inclusion criteria</td>
<td>clinically isolated syndrome or definite multiple sclerosis</td>
</tr>
<tr>
<td>treatment strategies</td>
<td>patients may contribute 6-month periods to either the treated (where exposed to disease modifying therapy for ≥15 days during the given period) or the untreated pseudo-cohort (where exposure to disease modifying therapy during the given period &lt;15 days)</td>
</tr>
<tr>
<td>assignment procedures</td>
<td>non-random assignation of therapy by treating neurologists</td>
</tr>
<tr>
<td>follow-up period</td>
<td>follow-up ≥1 year, ≥3 disability scores with ≥1 score recorded per year</td>
</tr>
<tr>
<td>outcomes</td>
<td>• 12-month confirmed disability worsening events (increase in EDSS by 1 step; 1.5 step if baseline EDSS=0 and 0.5 steps if baseline EDSS&gt;5.5)</td>
</tr>
<tr>
<td></td>
<td>• 12-month confirmed disability improvement events (decrease in EDSS by 1 step; 1.5 steps if baseline EDSS≤1.5 and 0.5 steps if baseline EDSS&gt;6)</td>
</tr>
<tr>
<td></td>
<td>• relapses</td>
</tr>
<tr>
<td>causal contrast of interest</td>
<td>per-protocol effect</td>
</tr>
<tr>
<td>analysis</td>
<td>proportional hazards models of multiple events with robust estimation of variance and inverse probability of treatment weights to adjust for fixed and time-dependent confounders and intermediates of outcomes; this analysis plan implies that data on the adjustment factors are available</td>
</tr>
</tbody>
</table>
Supplementary figure 1. An example of the included follow-up.

Only patients followed-up with prospective data entry commencing after 1st January 2006 were included. One treated / untreated epoch per patient was included. Patients were censored at the change of their treatment status.
**Supplementary figure 2.** Attrition of the number of patients over time for each therapy
**Supplementary Figure 3:** Covariate balance over time, the presented example shows the three reference treatments

**Supplementary figure 4:** Weighted and unweighted covariate balance measured with standardized difference between glatiramer acetate and alternative treatment groups in an ATT model.
Supplementary figure 5: Comparison of multiple therapies (ATE) among patients with recorded MRI data.