MSJ Experimental Translational Clinical

Original Research Paper

Comparative efficacy and discontinuation of dimethyl fumarate and fingolimod in clinical practice at 24-month follow-up

Carrie M Hersh, Thomas E Love, Anasua Bandyopadhyay, Samuel Cohn, Claire Hara-Cleaver, Robert A Bermel, Robert J Fox, Jeffrey A Cohen and Daniel Ontaneda

Abstract

Background: Dimethyl fumarate and fingolimod are oral disease-modifying therapies approved to treat relapsing multiple sclerosis. Prior observational studies and our previous 12-month investigation showed comparable clinical efficacy.

Objective: The purpose of this study was to assess real-world efficacy and discontinuation of dimethyl fumarate and fingolimod over 24 months in patients with multiple sclerosis.

Methods: Patients treated with dimethyl fumarate ($n = 395$) or fingolimod ($n = 264$) completed 24-month follow-up in a large academic multiple sclerosis center. Discontinuation rates and measures of disease activity were compared after propensity score weighting. The primary outcome was on-treatment annualized relapse rate ratio. Other measures included rate of drug discontinuation and brain magnetic resonance imaging activity defined as new T2 and/or gadolinium-enhancing lesions.

Results: Propensity score weighting showed excellent covariate balance. At 24 months, dimethyl fumarate demonstrated comparable annualized relapse rate (rate ratio = 1.45, 95% confidence interval 0.53−3.99) and brain magnetic resonance imaging activity (odds ratio = 1.38, 95% confidence interval 0.83−2.32). Dimethyl fumarate patients discontinued therapy earlier compared to fingolimod (hazard ratio = 1.40, 95% confidence interval 1.11−1.77) and were more likely to discontinue therapy due to intolerability (odds ratio = 1.98, 95% confidence interval 1.18−3.23).

Conclusion: Dimethyl fumarate and fingolimod had similar reductions in annualized relapse rate in clinical trials, and our real-world experience supports this observation. Dimethyl fumarate-treated patients had higher likelihood of early discontinuation, and this was mostly due to intolerability.

Introduction

Dimethyl fumarate (DMF) and fingolimod (FTY) are oral disease-modifying therapies (DMTs) commonly used in clinical practice to treat relapsing forms of multiple sclerosis (MS). On the basis of phase 3 clinical trials, DMF (240 mg, one tablet twice daily) was approved by the US Food and Drug Administration (FDA) and the European Medicines Agency in 2013.[1,2] FTY (0.5 mg, one tablet daily) was approved by the FDA in 2010 and the European Medicines Agency in 2011.[3,4] The DMTs showed similar efficacy in phase 3 randomized clinical trials (RCTs)[5] including reductions in annualized relapse rates (ARRs) and magnetic resonance imaging (MRI) disease activity. By 24 months, 7.5% of FTY patients in phase 3 clinical trials discontinued therapy due to adverse effects, while 14% of DMF patients discontinued treatment in respective RCTs due to intolerability.[1,2]

While RCTs are the standard approach for regulatory approval of new therapeutic agents, observational

Correspondence to:
Carrie M Hersh
Lou Ruvo Center for Brain Health, Cleveland Clinic, 888 W Bonneville Ave, Las Vegas, NV, 89106, USA.
hershc@ccf.org

Carrie M Hersh,
Lou Ruvo Center for Brain Health, Cleveland Clinic, USA

Thomas E Love,
Departments of Medicine
and Epidemiology and
Biostatistics, Case Western
Reserve University, USA

Anasua Bandyopadhyay,
College of Arts and
Sciences, Emory University,
USA

Samuel Cohn,
Department of Neurology,
Cleveland Clinic, USA

Claire Hara-Cleaver,
Mellen Center for Multiple
Sclerosis Treatment and
Research, Cleveland Clinic,
USA

Robert A Bermel,
Mellen Center for Multiple
Sclerosis Treatment and
Research, Cleveland Clinic,
USA

Robert J Fox,
Mellen Center for Multiple
Sclerosis Treatment and
Research, Cleveland Clinic,
USA

Jeffrey A Cohen,
Mellen Center for Multiple
Sclerosis Treatment and
Research, Cleveland Clinic,
USA

Daniel Ontaneda,
Mellen Center for Multiple
Sclerosis Treatment and
Research, Cleveland Clinic,
USA

studies are valuable when there are multiple available treatments and comparative RCTs are not feasible.[6] Observational studies also allow assessment of larger numbers of patients and may better reflect treatment effects in a real-world clinical setting.[6] However, observational studies are prone to confounding and bias, and approaches are needed to limit baseline imbalances in disease characteristics before direct comparisons are made. Propensity score (PS) analysis is a statistical method that reduces the impact of confounding and selection and indication biases by balancing the distributions of covariates across treatment groups, approximating a randomized study design.[7,8] Covariates are included in the PS model if they affect treatment decisions in relation to the therapeutics being studied. Standard PS methods for observational study analyses, such as matching and weighting on the PS, allow investigators to compare treatment groups in a manner that reduces bias.[9]

While separate clinical trials showed similar efficacy of DMF and FTY, there are limited head-to-head studies comparing these two agents to inform decision-making in routine practice. We sought to fill this knowledge gap by exploring comparative efficacy and discontinuation of DMF and FTY in a large MS tertiary referral center. Our previous 12-month study[10] demonstrated similar ARRs, neurologic measures of disability, overall brain MRI activity, and discontinuation between DMF and FTY. However, DMF-treated patients had higher proportion of gadolinium-enhancing (GdE) lesions and discontinued treatment earlier than FTY due to intolerability. The current investigation compares efficacy and discontinuation in patients treated with DMF or FTY in clinical practice over 24 months.

Materials and methods

Patient population
We conducted a retrospective observational study of patients with MS followed at the Cleveland Clinic Mellen Center and treated either with DMF or FTY with 24-month follow-up data available. The rationale for patient selection is described in our prior 12-month study.[10] Patients with progressive forms of MS were included in this study to reflect our real-world experience of DMT use in clinical practice. Subgroup analysis with only relapsing–remitting MS (RRMS) patients was also conducted.

Data collection
Following institutional review board approval, all patients prescribed DMF or FTY during the

designated time periods were identified. Review of the electronic medical record (EMR) was conducted to collect baseline and outcomes data 12 and 24 months after drug initiation. We incorporated baseline covariates derived from information collected from the EMR in the 12 months prior to respective treatment exposure that was verified by an MS specialty provider prior to inclusion.[11]

Clinical, radiographic, and patient-reported outcome (PRO) measures were collected from the EMR. Post-baseline follow-up (e.g. visit/MRI frequency and protocols) did not differ between the treatment groups. Clinical relapses, defined as new or worsening MS symptoms lasting greater than 24 h without a concomitant illness, were diagnosed by the treating clinicians. The timed 25-foot walk (T25FW; a quantitative measure of walking speed)[12] and nine-hole peg test (9-HPT; a quantitative measure of dexterity)[13] were measured by the treating clinicians as part of routine clinical practice. Number of new T2-hyperintense brain lesions and semi-quantitative assessment of overall lesion burden were manually counted and categorized, respectively, by the author (CH) and Cleveland Clinic neuroradiologists. Neuroradiological data at 12 months were compared to MRI 6–12 months prior to baseline MRI, and 24-month neuroradiological data were compared to on-treatment MRI within 12 months of drug initiation. PRO measures including Patient Health Questionnaire-9 (PHQ-9; a validated PRO scale for depression),[14] Multiple Sclerosis Performance Scale (MSPS; assessment of vision, hand function, sensation, spasticity, mobility, fatigue, cognition, and bladder and bowel control),[15] and European Quality of Life-5 Dimensions (EQ5D) quality of life scale[16] were collected.

We used REDCap software to create an encrypted database on secure Cleveland Clinic servers.

Statistical analysis
Data were imported for analysis into R version 3.3.1.[17] All covariates included in the PS model were missing in fewer than 10% of subjects. We accounted for missingness patterns in the PS model using the same approach as our previous 12-month study.[10] Covariates with proportion of missing information >10% were excluded from the model.

Analyses were conducted with patients on their respective treatments at 24-month follow-up. The primary outcome was ARR ratio (DMF vs FTY). The ARR was calculated by dividing the total number of relapses by the total number of

person-years at risk. Secondary outcomes included: time to first relapse; proportion who discontinued therapy; time to discontinuation; proportion of patients with new T2 lesions and/or GdE lesions; proportion of patients with 20% worsening on T25FW and 9-HPT; and proportion of patients with depression defined as PHQ-9 score ≥ 10 (indicating moderate depression or worse). We also explored the proportions of patients without disease activity, defined as absence of clinical relapses and brain MRI activity (composite measure of new T2 and/or GdE lesions).

PS weighting methods were as previously described.[10] The PS model was built as a logistic regression to calculate the likelihood of DMF initiation, as compared to FTY, using pre-determined demographics and baseline disease characteristics. A PS was derived for each patient and subsequently used in average treatment effect on the treated (ATT) weighting[18,19] to identify samples of patients in the DMF and FTY groups who were similar at baseline except for treatment.

The strength of the weighting technique was evaluated by how effectively it balanced the two treatment groups, as determined by comparing standardized differences in covariates before and after propensity adjustment. Excellent covariate balance was defined as an absolute standardized difference $<10\%$ on the means of the covariate across the two therapies. Following ATT weighting, treatment groups were subsequently compared using conditional logistic regression to measure odds ratio estimates for binary outcomes and Cox-proportional hazards model and Kaplan-Meier survival curves to obtain estimates for survival outcomes. On-treatment ARR was analyzed using a Poisson regression model. Odds ratios were calculated as patients treated with DMF compared to patients treated with FTY. The main outcome (ARR ratio) was based on one two-tailed test of statistical significance with $\alpha = 0.05$. Assuming 80% power with a total sample size = 659 patients, the minimum detectable effect size (MDES) was determined to be 0.23.

Results

Baseline characteristics
In the original cohort,[10] 12-month follow-up data were available for 775 patients (DMF $n = 458$, FTY $n = 317$). In the present study, 24-month follow-up data were available for 659 patients: DMF $n = 395$ (14% lost to follow-up) and FTY $n = 264$ (17% lost to follow-up). Baseline demographic and disease characteristics are presented in Table 1. There was a sizeable proportion of patients with progressive MS in our cohort (DMF = 25.8%, FTY = 18.6%).

Propensity model
The propensity model was built using covariates listed in Table 1. Missing data among covariates in the PS model did not meaningfully change the overall covariate balance following PS analysis. The model appropriately assigned higher PS to DMF vs FTY, and there was adequate overlap of PS between the two treatment groups (Figure 1). Before PS adjustment, the treatment groups were not well balanced with absolute value of the standardized difference of the linear PS comparing DMF to FTY = 87.8%, considerably over the 50% standard proposed by Rubin.[9] Thus, we favored the need for PS adjustment over unadjusted comparisons to account for observed selection bias.

ATT weighting effectively balanced treatment groups with only five covariates having absolute standardized differences greater than 10% (Figure 2). The weighting approach also yielded a similar linear PS distribution with a standardized difference of 17.8%, well within the 50% standard.

Outcome estimates for the overall cohort
A summary of unadjusted outcomes and post-weighting outcome estimates is presented in Tables 2 and 3.

By 24 months, 16.8% of DMF patients experienced a clinical relapse vs 17.0% of FTY patients. DMF patients experienced 66 relapses during 777.9 patient-years of treatment with ARR = 0.08 (95% confidence interval (CI) 0.30−0.46). FTY patients experienced 45 relapses during 552.7 patient-years of treatment with ARR = 0.08 (95% CI 0.33−0.48). After PS adjustment, there was no significant difference in ARR (rate ratio = 1.45, 95% CI 0.53−3.99). Mean time to first relapse was 3.83 months for DMF vs 7.56 months for FTY, though no significant difference was found in time to first relapse between groups over the first 24 months of treatment (hazard ratio (HR) = 1.26, 95% CI 0.87−1.81) (Figure 3(a)). Clinical disability measures were comparable between treatment groups after PS adjustment (Table 3).

By 24 months, 163 patients (41.3%) discontinued DMF and 94 patients (35.6%) discontinued FTY. Patients discontinued treatment most often due to intolerability (DMF $n = 100$, 25.3%; FTY $n = 39$,

Table 1. Baseline characteristics of dimethyl fumarate (DMF) and fingolimod (FTY) patient cohorts.

	DMF		FTY		
	n = 395		**n = 264**		
	n or mean	% or SD	*n* or mean	% or SD	*p*-Value
Age (years, SD)[a]	**46.9**	**11**	**44.1**	**9.2**	**0.001**
Female[a]	278	70.8%	187	70.4%	0.970
Race[a]					**0.008**
White	**333**	**84.3%**	243	92.0%	
Black	**44**	**11.1%**	18	6.8%	
Other	**18**	**4.6%**	3	1.1%	
Disease duration (years, SD)[a]	**14.5**	**9.0**	**16.3**	**7.9**	**0.007**
Relapsing–remitting MS[a]	**293**	**74.2%**	**215**	**81.4%**	**0.038**
Number of relapses[a]					0.122
0	237	61.9%	159	60.7%	
1	108	28.2%	90	34.4%	
2	26	6.8%	11	4.2%	
≥3	12	3.2%	2	0.8%	
DMF/FTY used as first-line agent[a]	39	7.8%	12	4.5%	0.139
Direct switch from prior DMT[a]	294	71.5%	197	73.5%	0.642
Reason for switch from prior DMT					
Clinical relapse[a]	**51**	**13.2%**	**51**	**19.3%**	**0.046**
MRI activity[a]	52	13.5%	30	11.4%	0.500
Disability progression[a]	**69**	**17.9%**	**73**	**27.7%**	**0.004**
Intolerance[a]	185	46.8%	125	47.3%	0.960
Cost/insurance coverage[a]	12	3.1%	8	3.0%	1.000
Last therapy prior to DMF or FTY[a]					**<0.001**
Interferon	**110**	**28.5%**	**96**	**36.4%**	
Glatiramer	**125**	**32.4%**	**76**	**28.8%**	
Natalizumab	**49**	**12.7%**	**41**	**15.5%**	
Immunosupression	**27**	**7.0%**	**28**	**10.6%**	
Number of prior DMTs (*n*, SD)[a]	2.0	1.4	2.2	1.4	0.172
Interferon[a]	**259**	**66.9%**	**204**	**77.3%**	**0.006**
Glatiramer[a]	201	51.9%	140	53.0%	0.846
Natalizumab[a]	82	21.2%	59	22.3%	0.798
Immunosupression[a]	**65**	**16.8%**	**63**	**23.9%**	**0.033**
Laboratory values					
Mean WBC ($\times 10^9$/l)	7.2	2.9	7.3	3.1	0.719
Mean ALC ($\times 10^9$/l)	2.0	1.6	3.7	2.4	0.207
MRI available for review	384	97.2%	261	98.9%	
Disease burden on MRI[a]					**<0.001**
Mild	**204**	**53.1%**	**93**	**35.6%**	
Moderate	**151**	**39.3%**	**134**	**51.3%**	
Severe	**29**	**7.6%**	**34**	**13.0%**	
Gadolinium enhancement on MRI[a]	89	23.2%	58	22.5%	0.467
New T2 lesions on MRI[a]	102	26.6%	52	20.2%	0.081
Objective measures					
T25FW (s, SD)[a]	7.4	8.2	8.2	9.9	0.245
Ambulation assistance[a]					0.390
None	299	80.2%	199	80.6%	
Unilateral	28	7.5%	26	10.5%	
Bilateral	45	12.1%	22	8.9%	

Table 1. Continued

| | DMF | | FTY | | |
| | n = 395 | | n = 264 | | |
	n or mean	% or SD	*n* or mean	% or SD	*p*-Value
9-HPT (s, SD)	25.8	12.8	26.1	12.7	0.795
Patient reported outcomes					
MSPS score (mean, SD)[a]	11.5	7.5	11.6	7.3	0.887
EQ5D score (mean, SD)[a]	720.2	204.9	733.7	189.0	0.446
PHQ-9 score (mean, SD)[a]	6.8	5.8	6.5	5.7	0.573

9-HPT: nine-hole peg test; ALC: absolute lymphocyte count; DMT: disease-modifying therapy; EQ5D: European Quality of Life-5 Dimensions; MRI: magnetic resonance imaging; MS: multiple sclerosis; MSPS: Multiple Sclerosis Performance Scale; PHQ9: Patient Health Questionnaire-9; PS: propensity score; SD: standard deviation; T25FW: timed 25-foot walk; WBC: white blood cell.
The pre-baseline period over which the number of relapses was recorded was 12 months.
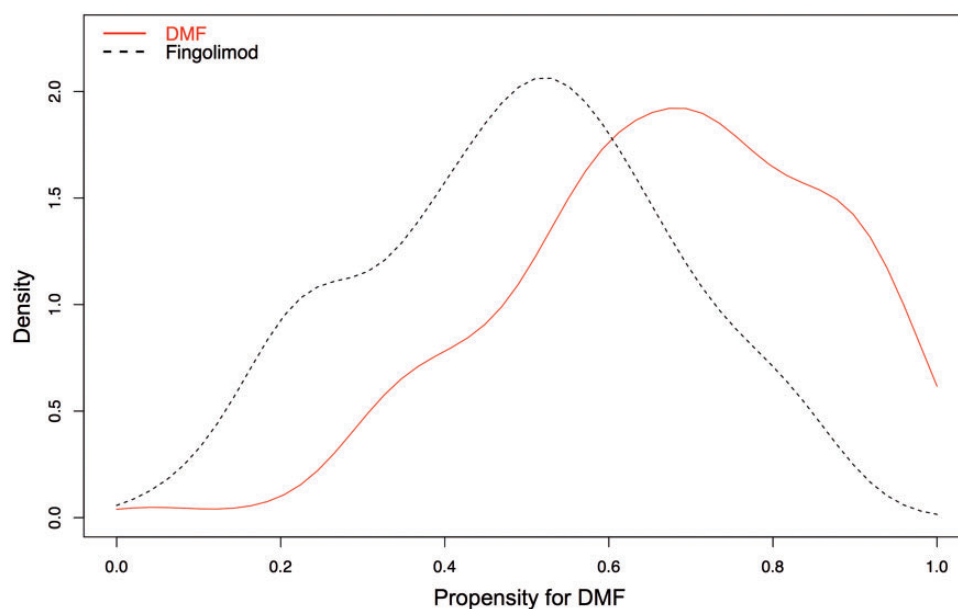[a]Covariates used in the PS model.

Figure 1. Density plot of propensity scores (propensity for dimethyl fumarate (DMF)) showing adequate overlap of propensity scores between DMF and fingolimod. MRI: magnetic resonance imaging; FTY: fingolimod.

14.8%). There was no significant difference in the proportion of patients who discontinued therapy between treatment groups (odds ratio (OR) = 1.42, 95% CI 0.94–2.14), though DMF patients were more likely to discontinue therapy due to intolerability compared to FTY patients (OR = 1.98, 95% CI 1.18–3.23). DMF patients also discontinued drug earlier than FTY patients (HR = 1.40, 95% CI 1.11–1.77), with mean time to discontinuation on DMF = 3.9 months vs FTY = 6.6 months. Kaplan-Meier survival curves confirmed earlier time to discontinuation in the DMF group (Figure 3(b)).

Patients treated with DMF had higher likelihood of depression at 24-month follow-up compared to those on FTY (OR = 2.18, 95% CI 1.14–4.15) with mean PHQ-9 score in DMF = 6.44 vs FTY = 5.26.

Among patients undergoing brain MRI while on DMT by 24-month follow-up, which included
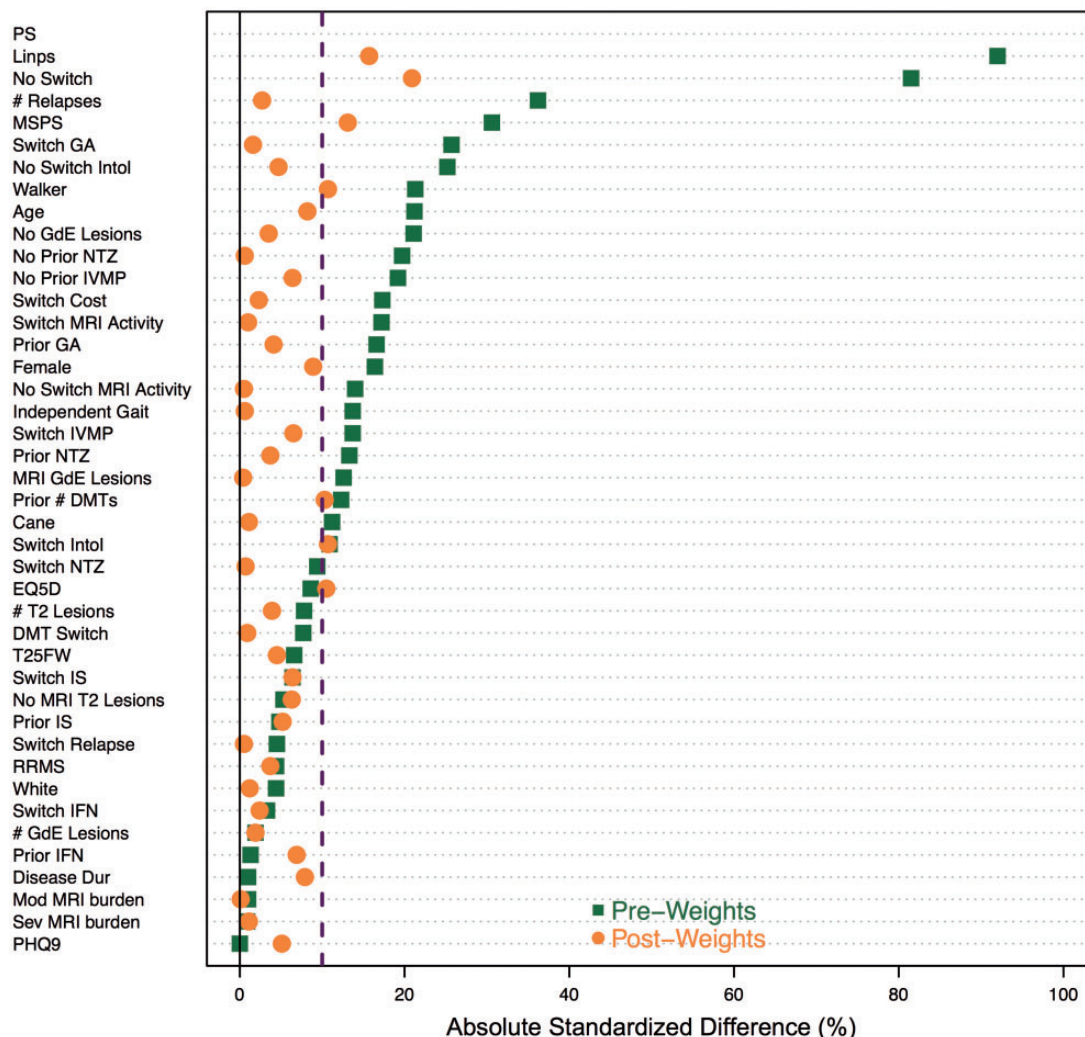
Figure 2. Absolute standardized difference plot comparing baseline covariates between dimethyl fumarate (DMF) and fingolimod (FTY) before and after average treatment effect on the treated (ATT) weighting using the linear propensity score.

Positive values represent higher standardized effect sizes for dimethyl fumarate (DMF).

DMT: disease-modifying therapy; EQ5D: European Quality of Life-5 Dimensions; GA: glatiramer acetate; GdE: gadolinium-enhancing; IFN: interferon; intol: intolerability; IS: immunosuppressive therapy; IVMP: intravenous methylprednisolone; linps: linear propensity score; MRI: magnetic resonance imaging; MS: multiple sclerosis; MSPS: Multiple Sclerosis Performance Scale; NTZ: natalizumab; PHQ9: Patient Health Questionnaire-9; PS: propensity score; T25FW: timed 25-foot walk.

available MRI data between 0–12 months and 12–24 months (DMF $n = 329$, FTY $n = 231$), 34.9% of patients on DMF demonstrated MRI activity (29.3% new T2 lesions), compared to 30.1% on FTY (21.5% new T2 lesions), but these differences were not statistically significant ($p = 0.20$). A sensitivity analysis comparing only patients with complete MRI data showed similar findings (data not shown). Cumulatively, there was greater risk of new GdE lesions in the DMF group compared to

FTY when accounting for brain MRI data collected by 24-month follow-up (OR = 2.07, 95% CI 1.04–4.12). However, DMF and FTY patients were equally likely to develop GdE lesions when only accounting for brain MRI between 12–24 months (OR = 1.64, 95% CI 0.82–5.63).

There was equal likelihood of absence of disease activity between DMF (63.1%) and FTY (66.7%) patients (OR = 1.31, 95% CI (0.80–2.13)) who had

Table 2. Summary of unadjusted outcomes at 24-month follow-up.

	DMF		Fingolimod		
	n = 395		*n* = 264		
	n	% or SD	*n*	% or SD	*p*-Value
Discontinued drug by 24 months	163	41.3%	94	35.6%	0.168
Discontinued drug between 12 and 24 months	42	10.6%	34	12.9%	0.080
Disease activity	40	10.1%	32	12.1%	0.498
Clinical relapse	9	2.3%	6	2.3%	
MRI activity	20	5.1%	9	3.4%	
Disability progression	12	3.0%	20	7.6%	
Intolerance/adverse effects	**100**	**25.3%**	**39**	**14.8%**	**0.002**
Lymphopenia	11	2.8%	3	1.1%	
Mean time to discontinuation (months, SD)	**3.90**	**3.54**	**6.58**	**4.23**	**<0.001**
Median time to discontinuation	3.0		6.0		
Clinical relapse by 24 months	66	16.8%	45	17.0%	1.000
Relapses per patient (mean, SD)	0.21	0.54	0.20	0.47	0.690
Mean time to first relapse (months, SD)	**3.83**	**2.78**	**7.56**	**4.34**	**<0.001**
MRI available for review by 24 months on DMT[a]	329	83.3%	231	87.5%	
Disease activity on MRI by 24 months on DMT	80	34.9%	49	30.1%	0.449
Gadolinium enhancement	34	14.8%	22	13.5%	0.360
New T2 lesions	67	29.3%	35	21.5%	0.250
MRI available for review at 24 months on DMT	167	42.3%	127	48.1%	
Disease activity on MRI between 12–24 months on DMT	31	18.6%	16	12.7%	
Gadolinium enhancement	19	11.4%	9	7.1%	0.298
New T2 lesions	23	13.8%	11	8.7%	0.250
Adverse effects (number of patients)	158	40.0%	108	40.9%	
Mean WBC ($\times10^9$/l)	**5.77**	**1.81**	**4.68**	**1.62**	**<0.001**
Mean ALC ($\times10^9$/l)	**1.34**	**0.98**	**0.55**	**0.32**	**<0.001**
Measures of neurologic disability					
T25FW (mean sec, SD)	**8.83 (n = 194)**	**8.78**	**6.96 (n = 159)**	**4.42**	**0.013**
20% worsening of T25FW	**82**	**42.3%**	**51**	**32.1%**	**0.049**
9-HPT- dominant (mean, SD)	24.81 (*n* = 111)	9.44	24.46 (*n* = 85)	9.96	0.766
20% worsening of 9 HPT- dominant	23	20.7%	21	24.7%	0.600
9-HPT- non-dominant (mean, SD)	27.52 (*n* = 133)	15.44	27.40 (*n* = 83)	15.97	0.949
20% worsening of 9 HPT- non-dominant	33	24.8%	20	24.1%	0.905
Patient reported outcomes					
PHQ-9 score (mean, SD)	**6.44 (n = 194)**	**5.43**	**5.26 (n = 155)**	**4.85**	**0.035**
PHQ-9 score ≥10	**54**	**27.8%**	**27**	**17.4%**	**0.030**

9-HPT: nine-hole peg test; ALC: absolute lymphocyte count; DMF: dimethyl fumarate; DMT: disease-modifying therapy; EQ5D: European Quality of Life-5 Dimensions; MRI: magnetic resonance imaging; MS: multiple sclerosis; MSPS: Multiple Sclerosis Performance Scale; PHQ9: Patient Health Questionnaire-9; SD: standard deviation; T25FW: timed 25-foot walk; WBC: white blood cell.
[a]cumulative MRI data that include 12-month and 24-month MRI time points.

Table 3. Unadjusted and adjusted discontinuation and efficacy outcomes at 24-month follow-up.

Discontinuation outcomes at 24 months		
Causal effect of treatment	Unadjusted	ATT weighting
Discontinuation	**1.27**	**1.42**
OR (95% CI)	(0.92–1.75)	(0.94–2.14)
Intolerability	**1.96[a]**	**1.98[a]**
OR (95% CI)	(1.30–2.94)	(1.18–3.23)
Breakthrough disease	**0.82**	**0.98**
OR (95% CI)	(0.50–1.34)	(0.54–1.79)
Time to discontinuation	**1.29[b]**	**1.40[a]**
Relative hazard ratio (95% CI)	(1.00–1.66)	(1.11–1.77)
Efficacy outcomes at 24 months: clinical measures of disease activity		
Causal effect of treatment	Unadjusted	ATT weighting
Annualized relapse rate (ARR)	**1.08**	**1.45**
ARR ratio (95% CI)	(0.74–1.58)	(0.53–3.99)
Proportion with relapses	**0.98**	**1.27**
OR (95% CI)	(0.65–1.49)	(0.72–2.23)
Time to first relapse	**1.02**	**1.26**
Relative hazard ratio (95% CI)	(0.70–1.49)	(0.87–1.81)
T25FW 20% worsening	**1.55[b]**	**1.23**
OR (95% CI)	(1.00–2.40)	(0.72–2.13)
9-Hole peg test 20% worsening	**0.80**	**0.79**
OR (95% CI)	(0.41–1.56)	(0.34–1.80)
Efficacy outcomes at 24 months: MRI measures of disease activity		
Causal effect of treatment	Unadjusted	ATT weighting
Brain MRI activity on DMT by 24 months[c]	**1.19**	**1.38**
OR (95% CI)	(0.80–1.79)	(0.83–2.32)
Brain MRI Gad-enhancing lesions	**1.36**	**2.07[b]**
OR (95% CI)	(0.77–2.39)	(1.04–4.12)
Brain MRI new T2 lesions	**1.43**	**1.38**
OR (95% CI)	(0.91–2.24)	(0.78–2.43)
Brain MRI activity on DMT at 24 months	**1.58**	**1.50**
OR (95% CI)	(0.82–3.04)	(0.90–4.37)
Brain MRI Gad-enhancing lesions	**1.68**	**1.64**
OR (95% CI)	(0.73–3.86)	(0.82–5.63)
Brain MRI new T2 lesions	**1.67**	**1.60**
OR (95% CI)	(0.78–3.57)	(0.80–5.12)
Efficacy outcomes at 24 months: patient reported outcome		
Causal effect of treatment	Unadjusted	ATT weighting
Depressed (PHQ9 ≥ 10)	**1.83[b]**	**2.18[b]**
OR (95% CI)	(1.09–3.08)	(1.14–4.15)

ATT: average treatment effect on the treated; CI: confidence interval; DMT: disease-modifying therapy; Gad: gadolinium; MRI: magnetic resonance imaging; OR: odds ratio; PHQ9: Patient Health Questionnaire-9; T25FW: timed 25-foot walk.
Significance level at alpha = 0.05.
[a]$p < 0.01$; [b]$p < 0.05$; [c]cumulative MRI data that include 12-month and 24-month MRI time points.
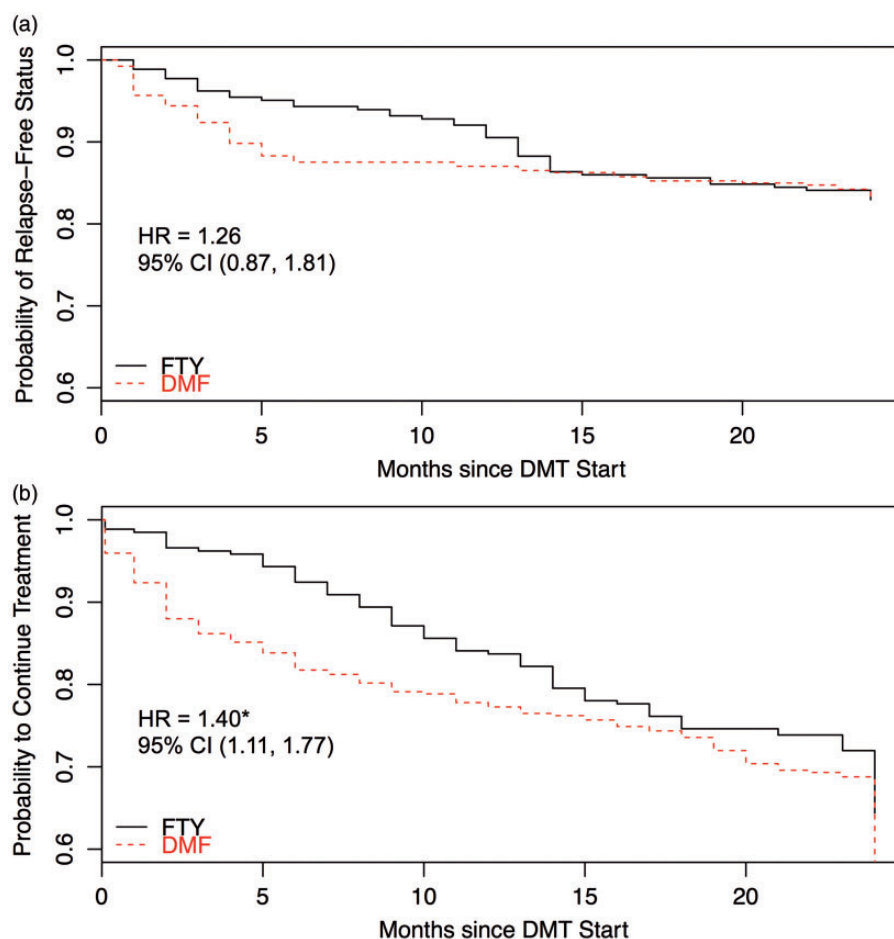
Figure 3. (a) Kaplan-Meier plot of relapse-free status through 24-month follow-up. (b) Kaplan-Meier plot of disease-modifying therapy (DMT) discontinuation through 24-month follow-up.
CI: confidence interval; DMF: dimethyl fumarate; FTY: fingolimod; HR: hazard ratio.

available on-treatment clinical and MRI data by 24 months (DMF $n = 271$, FTY $n = 231$).

Outcome estimates for the RRMS cohort

A summary of unadjusted outcomes and post-weighting outcome estimates for the RRMS subgroup is presented in Tables 4 and 5.

DMF RRMS patients experienced 52 relapses during 581.2 patient-years of treatment, with ARR = 0.09 (95% CI 0.31–0.52). FTY patients experienced 38 relapses during 552.7 patient-years of treatment with ARR = 0.08 (95% CI 0.32–0.46). PS adjusted subgroup analysis of RRMS patients on DMF vs FTY overall showed comparable findings to the larger cohort (Table 5); specifically, no significant difference in ARR, time to first relapse, brain MRI activity by 24 months, cumulative new T2 lesions, GdE lesions at 24 months, and 20% worsening on 9-HPT. RRMS patients on DMF were more likely to demonstrate 20% worsening on T25FW compared to FTY (OR = 2.06, 95% CI 1.15–3.70). The proportions of patients with PHQ-9 scores ≥ 10 were comparable between DMF and FTY (OR = 1.87, 95% CI 0.91–3.86).

DMF RRMS patients were more likely to discontinue treatment due to intolerability (OR = 2.29, 95% CI 1.28–4.09) with earlier time to discontinuation (HR = 1.71, 95% CI 1.29–2.26). However, in contrast to the larger group, the proportion of RRMS patients who discontinued DMF was significantly higher than those who discontinued FTY for any reason (OR = 1.85, 95% CI 1.16–2.95).

Among RRMS patients who had available on-treatment clinical and MRI data by 24 months (DMF $n = 202$, FTY $n = 189$), 59.4% of DMF patients ($n = 120$) and 64.6% of FTY patients ($n = 122$) showed absence of disease activity, also

Table 4. Summary of unadjusted outcomes for relapsing–remitting multiple sclerosis (RRMS) patients at 24-month follow-up.

	DMF		Fingolimod		
	n = 293		*n* = 215		
	n	% or SD	*n*	% or SD	*p*-Value
Discontinued drug by 24 months	**127**	**43.3%**	**70**	**32.6%**	**0.018**
Discontinued drug between 12 and 24 months	34	11.6%	22	10.2%	0.600
Disease activity	30	10.2%	23	10.7%	0.984
Clinical relapse	7	2.4%	3	1.4%	
MRI activity	14	4.8%	6	2.8%	
Disability progression	9	3.1%	14	6.5%	
Intolerance/adverse effects	**76**	**25.9%**	**31**	**14.4%**	**0.002**
Lymphopenia	2	0.7%	1	0.5%	
Mean time to discontinuation (months, SD)	**3.73**	**3.46**	**6.55**	**4.40**	**<0.001**
Median time to discontinuation					
Clinical relapse by 24 months (number of patients)	52	17.9%	38	17.7%	1.000
Relapses per patient (mean, SD)	0.22	0.51	0.20	0.46	0.719
Mean time to relapse (months, SD)	**3.95**	**3.04**	**8.09**	**4.22**	**<0.001**
MRI available for review by 24 months on DMT[a]	241	82.3%	188	87.4%	
Disease activity on MRI by 24 months on DMT	62	25.6%	44	23.3%	0.655
Gadolinium enhancement	26	12.9%	20	10.6%	0.586
New T2 lesions	50	20.7%	31	16.4%	0.318
MRI available for review at 24 months on DMT	119	40.6%	107	49.8%	
Disease activity on MRI between 12–24 months on DMT	20	16.8%	14	13.1%	0.552
Gadolinium enhancement	13	10.9%	8	7.5%	0.508
New T2 lesions	13	10.9%	9	8.5%	0.697
Adverse effects (number of patients)					
Mean WBC ($\times 10^9$/l)	**5.85**	**1.85**	**4.69**	**1.59**	**<0.001**
Mean ALC ($\times 10^9$/l)	**1.32**	**0.90**	**0.54**	**0.34**	**<0.001**
Measures of neurologic disability					
T25FW (mean s, SD)	**9.00 (*n* = 146)**	**9.76**	**6.86 (*n* = 139)**	**4.42**	**0.018**
20% worsening of T25FW	**71**	**48.6%**	**48**	**34.5%**	**0.022**
9-HPT- dominant (mean, SD)	24.82 (*n* = 83)	9.53	24.69 (*n* = 77)	10.49	0.921
20% worsening of 9-HPT- dominant	20	24.1%	21	27.3%	0.781
9-HPT- non-dominant (mean, SD)	27.05 (*n* = 98)	11.83	27.58 (*n* = 75)	16.86	0.781
20% worsening of 9 HPT- non-dominant	28	28.6%	20	26.7%	0.916
Patient reported outcomes					
PHQ-9 score (mean, SD)	6.15	5.34	5.45	4.95	0.267
PHQ-9 score ≥10	36	26.3%	25	18.9%	0.197

9-HPT: nine-hole peg test; ALC: absolute lymphocyte count; DMF: dimethyl fumarate; DMT: disease-modifying therapy; MRI: magnetic resonance imaging; PHQ9: Patient Health Questionnaire-9; SD: standard deviation; T25FW: timed 25-foot walk; WBC: white blood cell.
[a]cumulative MRI data that include 12-month and 24-month MRI time points.

Table 5. Unadjusted and adjusted discontinuation and efficacy outcomes for relapsing–remitting multiple sclerosis (RRMS) patients (total $n = 508$ (dimethyl fumarate (DMF) = 293, fingolimod (FTY) = 215)).

Discontinuation outcomes at 24 months		
Causal effect of treatment	Unadjusted	ATT weighting
Discontinuation	**1.58[a]**	**1.85[b]**
OR (95% CI)	(1.10–2.29)	(1.16–2.95)
Intolerability	**2.08[b]**	**2.29[b]**
OR (95% CI)	(1.31–3.30)	(1.28–4.09)
Breakthrough disease	**0.95**	**1.20**
OR (95% CI)	(0.54–1.69)	(0.60–2.42)
Time to discontinuation	**1.52[b]**	**1.71[c]**
Relative hazard ratio (95% CI)	(1.14–2.04)	(1.29–2.26)
Efficacy outcomes at 24 months: clinical measures of disease activity		
Causal effect of treatment	Unadjusted	ATT weighting
Annualized relapse rate (ARR)	**1.08**	**1.33**
ARR ratio (95% CI)	(0.72–1.62)	(0.52–3.43)
Proportion with relapses	**1.01**	**1.27**
OR (95% CI)	(0.64–1.61)	(0.70–2.31)
Time to first relapse	**1.05**	**1.27**
Relative hazard ratio (95% CI)	(0.69–1.59)	(0.84–1.92)
T25FW 20% worsening	**1.79[a]**	**2.06[a]**
OR (95% CI)	(1.11–2.89)	(1.15–3.70)
9-Hole peg test 20% worsening	**0.85**	**0.78**
OR (95% CI)	(0.42–1.72)	(0.34–1.79)
Efficacy outcomes at 24 months: MRI measures of disease activity		
Causal effect of treatment	Unadjusted	ATT weighting
Brain MRI activity on DMT by 24 months[d]	**1.14**	**1.37**
OR (95% CI)	(0.73–1.77)	(0.75–2.50)
Brain MRI Gad-enhancing lesions	**1.25**	**1.69**
OR (95% CI)	(0.67–2.32)	(0.75–3.84)
Brain MRI new T2 lesions	**1.33**	**1.48**
OR (95% CI)	(0.81–2.18)	(0.77–2.86)
Brain MRI activity on DMT at 24 months	**1.34**	**1.52**
OR (95% CI)	(0.64–2.81)	(0.59–3.93)
Brain MRI Gad-enhancing lesions	**1.52**	1.60
OR (95% CI)	(0.60–3.82)	(0.50–5.17)
Brain MRI new T2 lesions	**1.32**	**1.62**
OR (95% CI)	(0.54–3.23)	(0.52–5.03)
Efficacy outcomes at 24 months: patient reported outcome		
Causal effect of treatment	Unadjusted	ATT weighting
Depressed (PHQ9 ≥ 10)	**1.53**	**1.87**
OR (95% CI)	(0.86–2.72)	(0.91–3.86)

ATT: average treatment effect on the treated; CI: confidence interval; DMT: disease-modifying therapy; Gad: gadolinium; MRI: magnetic resonance imaging; OR: odds ratio; PHQ9: Patient Health Questionnaire-9; T25FW: timed 25-foot walk.
Significance level at alpha = 0.05.
[a]$p < 0.05$; [b]$p < 0.01$; [c]$p < 0.001$; [d]cumulative MRI data that include 12-month and 24-month MRI time points.

demonstrating similar proportions between groups as the larger cohort (OR = 1.51, 95% CI 0.89–2.57).

Discussion

The current study presents a comparison of efficacy and discontinuation of DMF and FTY over 24 months in clinical practice. Similar to our 12-month comparative efficacy study, we investigated the earliest clinical experience by focusing on the period immediately following FDA approval of each DMT. We were able to investigate a relatively large population from a single academic MS center using sample sizes similar to those of individual treatment arms in phase 3 trials,[1–4,20] despite having a proportion of patients lost to follow-up by 24 months from the original cohort.

DMF and FTY showed comparable clinical and radiographic efficacy, and both groups demonstrated low clinical disease activity, as measured by low ARRs throughout the 24-month study. This finding was consistent across the larger cohort and RRMS subgroup. Our results were similar to those seen in a 24-month PS-adjusted analysis comparing DMF- and FTY-treated patients in the pivotal clinical trials[21] and in a large claims database analysis.[22] Further, over the initial 24 months of treatment, both groups demonstrated comparable proportion of relapses (DMF = 16.8%, FTY = 17.0%) to those in the respective RCTs (DMF = 17.0–22.0%, FTY = 16.0–18.0%).[1–4] Cumulative brain MRI activity over 24 months was similar between DMF and FTY, although there was a consistent trend in increased likelihood of cumulative GdE lesions in DMF-treated patients by 24 months (including MRI at both the 12-month and 24-month assessments) compared to our 12-month experience. This treatment effect difference was lost with brain MRI data only at the 24-month time point; likely reflecting stabilization of inflammatory disease activity with longer DMT exposure, in addition to improved tolerability and therefore drug adherence with prolonged DMF treatment. The proportion of patients with GdE lesions at 24 months (DMF = 11.4%, FTY = 7.1%) was overall comparable to those in respective phase 3 clinical trials (DMF = 20%, FTY = 9.9%), supporting MRI efficacy in clinical practice. In this context, similar data in our 12- and 24-month investigations compared to those of prior observational studies and RCTs support comparable efficacy of both treatments in clinical practice. Further, these findings affirm consistent treatment effects in clinical practice with longer-term DMF and FTY exposure.

Our relapse rates were lower compared to phase 3 RCTs (DMF = 0.22, FTY = 0.18), demonstrating an overall low inflammatory profile of the entire cohort. This discrepancy may be explained by the inclusion of progressive patients in the present study and the fact that relapses may not be captured as precisely in clinical practice compared to RCTs.

Within the first three months of treatment in our cohort, a higher percentage of DMF patients discontinued therapy compared to FTY, largely driven by intolerability. However, a much smaller and similar percentage of DMF and FTY patients discontinued therapy in the second year of follow-up, which is reassuring for longer-term treatment management in clinical practice. DMF patients in the entire cohort showed higher patient-reported depression scores by 24-month follow-up, which might in part be explained by higher self-reported scores at baseline. While the magnitude of the difference in depression scores is small, it might explain, at least in part, the higher proportion with discontinuation in the DMF group.

Similar to our 12-month experience, the 24-month cohort showed higher rates of drug discontinuation due to adverse effects compared to RCTs, as similarly observed in another large single center observational study: 25% of DMF patients discontinued drug vs 12–16% in phase 3 trials and 15% of FTY patients discontinued drug vs 5.6–7.5% in phase 3 trials.[23] These differences emphasize the importance of investigating patients in a real-world setting to ascertain treatment effects that are likely more representative of the general population. There is an expectation of different motivations to continue treatment in routine practice vs a highly selected and motivated clinical trial population. Further, evaluation of our cohort afforded us the opportunity to highlight differences in discontinuation rates between the overall population and RRMS subgroup. We anticipate a relatively reduced willingness on the part of the provider and RRMS patient to continue a DMT with side-effects when other therapies are available. Patients with progressive MS in this cohort may have had different expectations, leading to continued DMT use.

Our 24-month RRMS data suggest a better treatment effect for FTY on disability as measured by the T25FW, though overall the differences are small and possibly not clinically meaningful given the comparable rates of walking disability amongst the overall group and equivalent 9-HPT outcomes in both the RRMS subgroup and larger cohort.

In response to the growing number of highly effective therapies, treatment expectations and goals have evolved to encompass potential remission from the symptoms of MS, known as freedom from disease activity or no evidence of disease activity (NEDA). We assessed absence of disease activity (relapses and MRI) to explore freedom of inflammatory disease in DMF and FTY. We compared our 24-month experience to those assessed in an indirect comparison model using clinical trial data from the respective therapies.[24] We observed similar proportions of DMF and FTY patients with absence of disease activity in both the larger and RRMS cohorts. This treatment effect differed from those findings reported by Nixon et al.[24] showing an increased relative risk of achieving NEDA status in FTY vs placebo than the estimated relative risk for DMF vs placebo in each respective trial population. We expect these distinctions from our experience are multifactorial: different population selections (real-world observational vs clinical trial data), definitions of NEDA (we did not collect Expanded Disability Status Scale (EDSS) scores), and type of comparisons made (direct vs indirect).

We utilized PS methods to improve the balance of baseline covariates between DMF and FTY. We interpret the data with caution, recognizing the possibility of residual biases (ascertainment and attrition biases), and inability to account for unmeasured covariates resulting in hidden bias. We believe the variables used in our PS model accurately reflect which baseline health characteristics are considered important in deciding between DMTs in clinical practice. Analysis of survival outcome measures, in addition to ARR ratio as the primary endpoint, allowed estimation of treatment effects that are more robust to attrition bias. Further, inclusion of a separate RRMS subgroup PS analysis allowed us to ascertain treatment effects in a population of patients similar to those in phase 3 clinical trials.

We recognize that there are several limitations in the present study, and our findings should be interpreted with caution, owing to the assumptions inherent in any PS modeling approach. The results may not be entirely representative of the general population since the investigation was completed in a single large academic MS center. We also included patients with progressive MS, and the overall low inflammatory activity profile may have obscured differences in treatment effects. However, incorporation of a RRMS subgroup analysis helped clarify our outcomes, and similar trends across this cohort compared to the larger group were reassuring. Missing MRI and PRO data were substantial, a common limitation of real world observational studies. Reassuringly, a sensitivity analysis of patients with complete data showed similar results.

Conclusions

These 24-month results largely confirm our previous 12-month study, indicating that there were no significant differences in effectiveness between patients treated with DMF or FTY in routine clinical practice, and patients treated with DMF were more likely to discontinue therapy early due to intolerability. While most outcome measures continued to favor FTY in the larger cohort and RRMS subgroup, for which some of these exploratory measures reached statistical significance, the magnitudes of these differences were small. These real-world comparative effectiveness data confirm the clinical and radiological efficacy of DMF and FTY in clinical practice and should assist clinicians in treatment decisions regarding these DMTs in the management of MS. To improve external validity, the authors are collaborating with another tertiary referral MS center in studying a larger cohort of patients to ascertain these treatment effects persist across multiple sites.

Conflicts of interest
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References
1. Fox R, Miller D, Phillips JT, et al. Placebo-controlled phase 3 study of oral BG-12 or glatiramer in multiple sclerosis. *N Engl J Med* 2012; 367: 1087–1997.
2. Gold R, Kappos L, Arnold DL, et al. Placebo-controlled phase 3 study of oral BG-12 for relapsing multiple sclerosis. *N Engl J Med* 2012; 367: 1098–1107.

3. Cohen JA, Barkhof F, Comi G, et al. Oral fingolimod or intramuscular interferon for relapsing multiple sclerosis. *N Engl J Med* 2010; 362: 402–415.

4. Kappos L, Radue E-W, O'Connor P, et al. A placebo-controlled trial of oral fingolimod in relapsing multiple sclerosis. *N Engl J Med* 2010; 362: 387–401.

5. Phillips J and Fox R. BG-12 in multiple sclerosis. *Semin Neurol* 2013; 33: 56–65.

6. Yang W, Zilov A, Soewondo P, et al. Observational studies: Going beyond the boundaries of randomized controlled trials. *Diabetes Res Clin Pract* 2010; 88: S3–S9.

7. D'Agostino R. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998; 17: 2265–2281.

8. Rosenbaum PR and Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; 70: 41–55.

9. Rubin DB. Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Serv Outcomes Res Methodol* 2001; 2: 169–188.

10. Hersh C, Love T, Cohn S, et al. Comparative efficacy and discontinuation of dimethyl fumarate and fingolimod in clinical practice at 12-month follow-up. *Mult Scler Relat Disord* 2016; 10: 44–52.

11. Katzan I, Speck M, Dopler C, et al. The Knowledge Program: An innovative, comprehensive, electronic data capture system and warehouse. *AMIA Annu Symp Proc* 2011; ▌▌: 683–692.

12. Rudick RA, Cutter G and Reingold S. The Multiple Sclerosis Functional Composite: A new clinical outcome measure for multiple sclerosis trials. *Mult Scler* 2002; 8: 359–365.

13. Kragt JJ, van der Linden FA, Nielsen JM, et al. Clinical impact of 20% worsening on timed 25-foot walk and 9-hole peg test in multiple sclerosis. *Mult Scler* 2006; 12: 594–598.

14. Kroenke K, Spitzer RL and Williams JB. The PHQ-9: Validity of a brief depression severity measure. *J Gen Intern Med* 2001; 16: 606–613.

15. Schwartz CE, Vollmer T and Lee H. Reliability and validity of two self-report measures of impairment and disability for MS. North American Research Consortium on Multiple Sclerosis Outcomes Study Group. *Neurology* 1999; 52: 63–70.

16. The EuroQol Group. EuroQol-a new facility for the measurement of health-related quality of life. *Health Policy* 1990; 16: 199–208.

17. Team RC. *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing, 2014.

18. Hirano K, Imbens GW and Ridder G. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* 2003; 71: 1161–1189.

19. Austin P. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* 2011; 46: 399–424.

20. Calabresi PA, Radue EW, Goodin DS, et al. Safety and efficacy of fingolimod in patients with relapsing–remitting multiple sclerosis (FREEDOMS II): A double-blind, randomised, placebo-controlled, phase 3 trial. *Lancet Neurol* 2014; 13: 545–556.

21. Fox R, Chan A, Zhang A, et al. Comparative effectiveness using a matching-adjusted indirect comparison between delayed-release dimethyl fumarate and fingolimod for the treatment of multiple sclerosis. *Curr Med Res Opin* 2017; 33: 175–183.

22. Boster A, Nicholas J, Wu N, et al. Comparative effectiveness research of disease-modifying therapies for the management of multiple sclerosis: Analysis of a large health insurance claims database. *Neurol Ther* 2017; 6: 91–102.

23. Vollmer B, Kavita N, Sillas S, et al. Comparison of fingolimod and dimethyl fumarate in the treatment of multiple sclerosis: Two year experience. *Neurology* 2016; S24.004. Available at: http://www.neurology.org/content/86/16_Supplement/S24.004.

24. Nixon R, Bergvall N, Tomic D, et al. No evidence of disease activity: Indirect comparisons of oral therapies for the treatment of relapsing–remitting multiple sclerosis. *Adv Ther* 2014; 31: 1134–1154.