Harmonization of real-world studies in multiple sclerosis: retrospective analysis from the RIReMS group

Marcello Moccia[1], Pietro Annovazzi[2], Maria Chiara Buscarinu[3], Massimiliano Calabrese[4], Paola Cavalla[5], Cinzia Cordioli[6], Massimiliano Di Filippo[7], Diana Ferraro[8], Alberto Gajofatto[4], Antonio Gallo[9], Roberta Lanzillo[1], Alice Laroni[10-11], Lorena Lorefice[12], Simona Mallucchi[13], Viviana Nociti[14], Damiano Paolicelli[15], Federica Pinardi[16], Luca Prosperini[17], Marta Radaelli[18], Paolo Ragonese[19], Valentina Tomassini[20-22], Carla Tortorella[17], Eleonora Cocco[12], Claudio Gasperini[17], and Claudio Solaro[23].

1. MS Clinical Care and Research Centre, Department of Neuroscience, Federico II University of Naples, Italy
2. MS Centre, ASST Valle Olona, Gallarate Hospital, Varese, Italy
3. Department of Neurosciences, Mental Health and Sensory Organs, Sapienza University, Rome, Italy
4. Department of Neuroscience, Biomedicine and Movement, University of Verona, Italy
5. MS Center, Department of Neurosciences and Mental Health, AOU City of Health & Science University Hospital, Turin, Italy
6. Multiple Sclerosis Center, ASST Spedali Civili di Brescia, Brescia, Italy
7. Neurology Unit, Medicine Department, University of Perugia, Italy
8. Department of Biomedical, Metabolic and Neurosciences, University of Modena and Reggio Emilia, Italy

9. Department of Advanced Medical and Surgical Sciences, University of Campania "L. Vanvitelli", Naples, Italy
10. Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health and Center of Excellence for Biomedical Research (CEBR), University of Genova, Italy
11. IRCCS Ospedale Policlinico San Martino, Genoa, Italy
12. Department of Medical Sciences and Public Health, University of Cagliari, Italy
13. Multiple Sclerosis Centre, San Luigi Gonzaga Hospital, Orbassano, Turin, Italy
14. Multiple Sclerosis Center, Fondazione Policlinico Universitario 'A. Gemelli' IRCCS, Università Cattolica del Sacro Cuore, Rome, Italy
15. Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of Bari, Italy
16. UOSI Multiple Sclerosis Rehabilitation, IRCCS, Bologna, Italy
17. Department of Neurosciences, Ospedale San Camillo Forlanini, Rome, Italy
18. Department of Neurology, San Raffaele Hospital, Milan, Italy
19. Department of Biomedicine Neurosciences and advanced Diagnostic (BiND), University of Palermo, Italy
20. Institute for Biomedical Technologies (ITAB), Department of Neurosciences, Imaging and Clinical Sciences, University of Chieti-Pescara “G. d’Annunzio”, Chieti, Italy
21. MS Centre, Neurology Unit, SS. Annunziata University Hospital, Chieti, Italy
22. Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University School of Medicine, United Kingdom
23. Rehabilitation Department, Mons. L. Novarese, Moncrivello, Vercelli, Italy

CORRESPONDENCE TO: Marcello Moccia

Multiple Sclerosis Clinical Care and Research Centre

Department of Neuroscience, Reproductive Science and Odontostomatology

Federico II University of Naples

Via Sergio Pansini 5, 80131 Naples, Italy

Tel./Fax +39 081 7462670

Email. marcello.moccia@unina.it moccia.marcello@gmail.com

Harmonization of real-world studies in multiple sclerosis: retrospective analysis from the RIReMS group

ABSTRACT

Background. Worldwide multiple sclerosis (MS) centers have coordinated their efforts to use data acquired in clinical practice for real-world observational studies. In this retrospective study, we aim to harmonize outcome measures, and to evaluate their heterogeneity within the Rising Italian Researchers in MS (RIReMS) study group.

Methods. RIReMS members filled in a structured questionnaire evaluating the use of different outcome measures in clinical practice. Thereafter, thirty-four already-published papers from RIReMS centers were used for heterogeneity analyses, using the DerSimonian and Laird random-effects method to compute the between-study variance (τ^2).

Results. Based on questionnaire results, we defined basic modules for diagnosis and follow-up, consisting of outcome measures recorded by all participating centers at the time of diagnosis, and, then, at least annually; we also defined more detailed/optional modules, with outcome measures recorded less frequently and/or in the presence of specific clinical indications. Looking at heterogeneity, we found 5-year variance in age at onset (ES=27.34; 95%CI=26.18, 28.49; p<0.01; τ^2=4.76), and 7% in female percent (ES=66.42; 95%CI=63.08, 69.76; p<0.01; τ^2=7.15). EDSS variance was 0.2 in studies including patients with average age <36.1 years (ES=1.96; 95%CI=1.69, 2.24; p<0.01; τ^2=0.19), or from 36.8 to 41.1 years (ES=2.70; 95%CI=2.39, 3.01; p<0.01; τ^2=0.18), but increased to 3 in studies including patients aged >41.4 years (ES=4.37; 95%CI=3.40, 5.35; p<0.01; τ^2=2.96). The lowest variance of relapse rate was found in studies with follow-up duration ≤2 years

(ES=9.07; 95%CI=5.21, 12.93; p=0.02; τ^2=5.53), whilst the lowest variance in EDSS progression was found in studies with follow-up duration >2 years (ES=5.41; 95%CI=3.22, 7.60; p=0.02; τ^2=1.00).

Discussion. We suggest common sets of biomarkers to be acquired in clinical practice, that can be used for research purposes. Also, we provide researchers with specific indications for improving inclusion criteria and data analysis, ultimately allowing data harmonization and high-quality collaborative studies.

KEY WORDS: multiple sclerosis; real world; harmonization; outcome measures.

INTRODUCTION

Multiple sclerosis (MS) is a chronic immuno-mediated disease of the central nervous system, characterised by young age at onset, female predominance, and heterogeneous symptoms (Thompson et al., 2018). As such, enormous effort has been put into the identification and characterisation of the variety of MS symptoms (e.g., motor, cognitive, sensory, etc) through disease progression, using clinical, cognitive, neuroimaging, laboratory, and neurophysiological outcome measures (Tur et al., 2018).

To overcome sample size constraints, worldwide MS centres have coordinated their efforts to use data acquired in clinical practice for real-world observational studies, mostly aiming at identifying predictors of poor outcome and treatment response/failure (Glaser et al., 2019). However, in the absence of a consensus, real-world studies generally include the limited number of variables recorded in clinical practice by all participating centres (Middleton et al., 2018; Trojano et al., 2017). As such, the harmonization of outcome measures collected in clinical practice could be helpful to produce more granular real-world studies. In a previous study, Salter and colleagues showed the feasibility of harmonization between North American Research Committee on MS (NARCOMS) Registry, German MS Register (GMSR), and United Kingdom MS (UK-MS) Register (Salter et al., 2020). However, MS registries and real-world studies are composed by data collected from different MS centres, that are not necessarily homogeneous (e.g., selection bias) (Fortier et al., 2017; Salter et al., 2020; Trojano et al., 2017).

Among collaborative research networks contributing to real world studies, the Rising Italian Researchers in MS (RIReMS) (www.rirems.it) study group is composed of experienced neurologists from representative Italian MS Centres, and has already promoted many cross-sectional and

longitudinal real-world studies (Ferraro et al., 2020a; Lanzillo et al., 2018). Thus, studying the extent to which populations across RIReMS sites are homogeneous could be helpful for future study design within RIReMS and other collaborative networks. In this retrospective study, we aim to: (1) evaluate preferred clinical, cognitive, neuroimaging, laboratory, and neurophysiological outcome measures in RIReMS centres; (2) define modular sets of biomarkers for the clinical practice and for research purposes in RIReMS centres; and (3) measure the heterogeneity of MS populations included in different studies from RIReMS centres.

METHODS

Study design

The present study was conducted within 20 RIReMS centres. In the first part of the study, RIReMS members were asked to fill in a structured questionnaire evaluating outcome measures used in clinical practice and/or for research purposes at their MS Centres, which were used to define modular sets of outcome measures. Then, clinical studies published by RIReMS centres were used for retrospective heterogeneity analyses. Results were discussed during a RIReMS meeting in January 2020 in advance of manuscript preparation.

Questionnaire for outcome measures

To define clinical, cognitive, neuroimaging, laboratory, and neurophysiological outcome measures (aim 1), each RIReMS member was required to fill in a structured questionnaire (English version in **Supplementary Material 1**). This questionnaire consisted of a list of demographics, clinical features (e.g., examination, scales, etc), cognitive tests, patient-reported outcome measures (PROMs), laboratory analyses, neurophysiological exams, and MRI parameters. For each outcome measure, participants were required to report: (1) whether they had been using them for their clinical practice and/or research purposes; (2) when and how frequently (e.g., diagnosis, follow-up, annually, in the presence of specific symptoms); (3) and their usefulness. Participants were also asked to share any additional thoughts in an open question.

Modular sets of outcome measures

Based on this questionnaire, modular sets of assessments were defined (aim 2). In particular, basic modules for diagnosis and for follow-up consisted of outcome measures recorded by all participating centres at the time of diagnosis, and, then, at least annually during follow-up. The

detailed module included outcome measures recorded in most participating centres in the presence of specific clinical indications. Finally, we also defined an optional extension module, including outcome measures recorded in some participating centres in the presence of specific clinical scenarios.

Heterogeneity analyses

Each participating RIReMS member was invited to suggest up to three previously-published papers with the following characteristics: 1) independent populations; 2) publication in the past 10 years; 3) inclusion of a wide range of outcome measures. Overall, participating RIReMS members suggested thirty-four papers, which were reviewed for consistency with inclusion criteria by two independent assessors and, then, were included in the analyses (**Table 1**) (Bajrami et al., 2018; Bisecco et al., 2018; Buscarinu et al., 2017; Calabrese et al., 2017; Camerota et al., 2017; Carotenuto et al., 2019; Cocozza et al., 2019; Coghe et al., 2018; De Biasi et al., 2016; Della Corte et al., 2018; Di Filippo et al., 2014; Di Gregorio et al., 2018; Ferraro et al., 2020b; Gaetani et al., 2019; Gajofatto et al., 2014; Lanzillo et al., 2017; Lipp et al., 2020; Lorefice et al., 2019; Mallucci et al., 2018; Malucchi et al., 2017; Manni et al., 2019; Marastoni et al., 2019; Megna et al., 2019; Moccia et al., 2019, 2016; Paolicelli et al., 2016; Pellegrino et al., 2018; Realmuto et al., 2019; Salemi et al., 2019; Scalfari et al., 2018; Sola et al., 2010; Vercellino et al., 2009; Zoccolella et al., 2012). From these studies, we extracted mean (and standard deviation), and rates of different outcome measures, as appropriate. Average age at onset, if not directly available, was calculated as the difference between age and disease duration reported in the study; standard deviation of the age was then included. If data was presented differently (e.g., median and range), conversion was performed with previously described methods (Wan et al., 2014). If the selected study was multicentre, data extraction and conversion only covered the sub-population from the RIReMS centre.

We preliminary summarised demographics and clinical features from different studies and, since no obvious differences were detected (e.g., outliers), all studies were included in subsequent analyses. To retrospectively evaluate heterogeneity in demographics and clinical features (aim 3), we estimated the between-study variance (τ^2) using the DerSimonian and Laird random-effects method, which includes studies' effect estimates and standard errors as input, as in a previous similar study (Salter et al., 2020); weights from these analyses were used for forest plot graphical presentation. We also estimated the percentage of between-study heterogeneity that is attributable to variability in the true treatment effect (I^2). As such, we obtained the amount of heterogeneity (τ^2), and the percent this heterogeneity was actually true, and not related to sampling variation (I^2) (Harris et al., 2008; Salter et al., 2020). These statistical models were applied to demographics and clinical features, and also to sub-analyses by subgroups (e.g., age, study duration). Results are presented as effect size (ES), 95% confidence intervals (95%CI), and p-value. Statistical significance was set at $p<0.05$. Statistical analyses were performed using Stata 15.0.

RESULTS

Outcome measures and modular sets of biomarkers

Based on the questionnaire for the use of different outcome measures, we defined four modular sets: basic module at diagnosis, basic module at follow-up, detailed module, and optional extension module. Full details of modular sets are reported in **Figure 1**.

Heterogeneity of RIReMS studies

Looking at demographics on RRMS patients, we found a variance of 5 years in age at onset (n=16) (ES=27.34; 95%CI=26.18, 28.49; p<0.01; τ^2=4.76; I^2=91.70%) (**Figure 2a**). We found a variance of 7% in the percent of females (n=18) (ES=66.42; 95%CI=63.08, 69.76; p<0.01; τ^2=7.15), only in part attributable to actual heterogeneity (I^2=51.70%) (**Figure 2b**).

Looking at motor disability on the whole population (n=34), we found a variance of 2.7 in expanded disability status scale (EDSS) (ES=3.08; 95%CI=2.50, 3.65; p<0.01; τ^2=2.73; I^2=99.70%), which decreased to 0.2 when considering studies including patients with average age <36.1 years (n=11) (ES=1.96; 95%CI=1.69, 2.24; p<0.01; τ^2=0.19; I^2=96.30%), or from 36.8 to 41.1 years (n=9) (ES=2.70; 95%CI=2.39, 3.01; p<0.01; τ^2=0.18; I^2=90.12%); on the contrary, EDSS variance increased to 3, when considering studies on more advanced disease stages (age>41.4 years) (n=12) (ES=4.37; 95%CI=3.40, 5.35; p<0.01; τ^2=2.96; I^2=99.70%) (**Figure 3**). Looking at more specific measures of motor disability, we found a variance of 41 seconds in time 25-feet walking test (n=5) (ES=15.19; 95%CI=9.31, 21.06; p<0.01; τ^2=41.61; I^2=97.21%), and of 27 seconds in 9-hole peg test (n=3) (ES=30.70; 95%CI=23.72, 37.68; p<0.01; τ^2=27.24; I^2=83.40%).

Looking at longitudinal studies, we found a variance of 19.4% in the yearly rate of patients with relapses (n=8) (ES=10.74; 95%CI=6.96, 14.53; p<0.01; τ^2=19.47; I^2=70.11%), which decreased to 5.5% in studies with follow-up duration ≤2 years (n=5) (ES=9.07; 95%CI=5.21, 12.93; p=0.02; τ^2=5.53; I^2=28.54%), and increased to 31.7% in studies with follow-up duration >2 years (n=3) (ES=12.26; 95%CI=5.41, 19.11; p<0.01; τ^2=31.72; I^2=86.65%) (**Figure 4a**). On the contrary, we found a variance of 5.8% in the yearly rate of patients with EDSS progression (n=6) (ES=4.81; 95%CI=2.21, 7.42; p=0.03; τ^2=5.83; I^2=59.07%), which increased to 15.6% in studies with follow-up duration ≤2 years (n=3) (ES=4.59; 95%CI=-0.74, 9.92; p=0.02; τ^2=15.68; I^2=71.90%), and decreased to 1% in studies with follow-up duration >2 years (n=3) (ES=5.41; 95%CI=3.22, 7.60; p=0.02; τ^2=1.00; I^2=0.00%) (**Figure 4b**).

Looking at MRI measures, we found a variance of 402 mL in brain volume (n=3) (ES=1354.30; 95%CI=1142.35, 1566.26; p=0.09; τ^2=402.67; I^2=57.60%), and 245 mL in grey matter volume (n=4) (ES=727.36; 95%CI=636.71, 818.00; p<0.01; τ^2=245.44; I^2=99.60%).

Looking at cognitive tests, we found a variance of 17 in symbol digit modalities test score (n=7) (ES=40.34; 95%CI=34.24, 46.44; p<0.01; τ^2=17.67; I^2=96.35%), and of 25 in paced auditory serial addition test (3 seconds) score (n=6) (ES=34.72; 95%CI=30.46, 38.98; p<0.01; τ^2=25.41; I^2=94.14%).

Looking at PROMs, we found a variance of 18 in Beck depression inventory (n=4) (ES=11.56; 95%CI=7.16, 15.96; p<0.01; τ^2=18.8; I^2=94.05%), of 39 in fatigue severity scale (n=4) (ES=40.59; 95%CI=31.39, 49.80; p<0.01; τ^2=39.70; I^2=95.10%), and of 6.5 in modified fatigue impact scale (n=4) (ES=38.84; 95%CI=33.57, 44.10; p=0.96; τ^2=0.0; I^2=0.0%).

DISCUSSION

In the present retrospective study, we compared the use of clinical, cognitive, neuroimaging, laboratory, and neurophysiological outcome measures between 20 RIReMS centres, and defined modular sets of biomarkers for the clinical practice and for research purposes. Also, we evaluated the variance of different outcome measures between RIReMS centres, and showed that these populations can be combined reliably, but with some caveats. As such, present results will be helpful for future study design, and will possibly improve harmonization of data collection in clinical practice.

We showed that all RIReMS centres have collected a number of basic outcome measures at the time of diagnosis, and, then, over the follow-up (at least annually), from patients seen routinely in outpatient clinics, irrespectively of whether they were currently recruited for a study. Our basic modules are in line with the current standard of clinical registries (Glaser et al., 2019; Pugliatti et al., 2012). Also, based on the experience of RIReMS centres, we developed detailed and optional extension modules with suggestions for using specific tools/scales to evaluate MS symptoms. Looking at outcome measures selected for different modules, we found a consensus about the functions/domains that should be assessed, but, less so, about the nature of specific assessments. This could be explained by differences in study designs, preferences based on previous experience, and/or availability of resources. However, our study also highlights areas of improvements for assessing specific symptoms. For instance, MS-specific scales for pain and sexual dysfunction are not validated yet in Italian language. Also, the collection of some variables (e.g., comorbidities) could be improved with specific classifications. Not least, outcome measures from detailed and optional extension modules might be biased by the collection depending on the clinical indications (e.g., ophthalmological evaluations in patients treated with fingolimod, spasticity scales in patients

treated with cannabinoids), and, thus, should be interpreted cautiously. Alternatively, basic modules could be expanded, with the inclusion of a wider range of outcome measures to be collected in clinical practice and, then, used in real-world studies. Of course, the number of participants, and the available staff (and costs) should be considered when extending the assessments to detailed and optional extension modules in clinical practice.

Thirty-four previously-published studies from RIReMS centres overall showed good homogeneity, but with some caveats. Age at onset presented with 5-year variance, mostly coming from the inclusion of studies conducted in different years; in particular, older age at onset was found in studies conducted in less recent years (Vercellino et al., 2009), or including historical cohorts (Moccia et al., 2016), when compared with contemporary populations, where earlier diagnosis is made possible by the application of newer diagnostic criteria (Brownlee et al., 2015; Petruzzo et al., 2020). On the contrary, the 7% variance of females is related to actual sampling variations, possibly as a consequence of studies with smaller sample size (Malucchi et al., 2017). Disability measured with the EDSS is known to be highly dependent on age (Manouchehrinia et al., 2017), and, accordingly, we found very small variance when including selected age ranges. The largest variability of disability was found in more advanced disease stages, when individual trajectories of progression have diverged (Bodini et al., 2020; Signori et al., 2018). Looking at longitudinal measures, variance was smaller within 2 years from study inclusion for relapses, and after 2 years for EDSS progression. These findings are in line with previous studies suggesting that, over time, the number of relapses progressively reduce (increasing population variability) (Schwehr et al., 2018), whilst disability outcomes need time to fully disclose (Kalincik et al., 2015). Of note, we did not perform a systematic review and meta-analysis of previously-published papers from RIReMS members (e.g., PRISMA checklist and flow diagram), which was out of the scope of the present manuscript, but relied on

participating members' suggestions for paper selection to depict heterogeneity of demographics and MS clinical features in different RIReMS research populations.

Based on the lesson learned from our previous experiences and from our retrospective heterogeneity analyses, we have integrated MS-specific recommendations to the Maelstrom Research guidelines for rigorous retrospective data harmonization (Fortier et al., 2017). The Maelstrom guidelines consist of a six-step checklist to achieve a successful harmonization, from dataset generation to result dissemination (Fortier et al., 2017). Here, we have developed MS-specific suggestions for each step of Maelstrom guidelines (**Figure 5**):

- **Step 0: define the questions and objectives**. Research questions and study objectives should be defined based on the availability of common outcome measures, which could be guided by our modules.
- **Step 1: assemble information and select studies**. Patients should be included within the same time period, or, at least, with diagnosis performed with same criteria. Age should be accounted for as a determinant of disability (e.g., age range in inclusion criteria, stratification of statistical analyses). Participating centres should contribute with populations of similar size, or, at least, specific statistical methods should be considered to reduce between-centre variability (e.g., propensity score matching).
- **Step 2: define variables and evaluate harmonization potential**. The study duration should be set depending on the primary outcomes (e.g., 2 years could be reasonable for relapses, but not enough for EDSS progression).
- **Step 3: process data**. Based on the previous steps, possible confounding should be evaluated, and, if necessary, should be considered statistically (e.g., age could be included as a covariate in the statistical models).

- **Step 4: estimate the quality of harmonized dataset(s) generated**. DerSimonian and Laird random-effects method could be used to evaluate the between-population variance in future studies.
- **Step 5: disseminate and preserve final harmonization products**. Harmonization analyses should be made available (e.g., as supplementary material).

Unfortunately, MRI measures, neuropsychological tests, and PROMs were included in a limited number of studies, and, thus, it is difficult to draw final conclusions. In the future, more detailed outcome measures could be used also for stratifying study populations, further improving the homogeneity. An additional limitation of this study is that we only analysed results from RIReMS centres, limiting results' generalisability, also considering the possible use of different tools/scales in other countries/languages. Not least, we did not evaluate heterogeneity in acquiring different outcome measures (e.g., methods for quantification of brain atrophy), which would need more in-depth evaluation in future studies. However, on this ground, further cooperation between different national and international centres should be implemented (Pugliatti et al., 2012). Also, we performed retrospective analyses on already-published data to reflect the general practice of clinical registries, where data is collected locally and then analysed retrospectively (Glaser et al., 2019; Trojano et al., 2019). In a previous study, Salter and colleagues investigated retrospective harmonization of three MS registries from three different areas (Germany, North America, and the UK), with ad-hoc data collection for employment status (Salter et al., 2020); by contrast, we investigated heterogeneity at the level of MS centres (which then contribute to formation of MS registries), by evaluating common sets of demographics and clinical features, cross-sectionally and longitudinally.

In conclusion, we suggested modular sets of biomarkers for data collection in clinical practice and for research purposes, that, if acquired by multiple MS centres, could improve data harmonization and sharing. Also, by analysing the heterogeneity of previous RIReMS papers retrospectively, we developed specific indications for future study design and data analysis. Common sets of biomarkers, along with improved study design and data analysis, will increase our chance to answer, more reliably, complex research questions with real-word data, ultimately accelerating MS research.

REFERENCES

Bajrami, A., Pitteri, M., Castellaro, M., Pizzini, F., Romualdi, C., Montemezzi, S., Monaco, S., Calabrese, M., 2018. The effect of fingolimod on focal and diffuse grey matter damage in active MS patients. J Neurol. 265, 2154–2161. doi:10.1007/s00415-018-8952-2

Bisecco, A., Stamenova, S., Caiazzo, G., D'Ambrosio, A., Sacco, R., Docimo, R., Esposito, S., Cirillo, M., Esposito, F., Bonavita, S., Tedeschi, G., Gallo, A., 2018. Attention and processing speed performance in multiple sclerosis is mostly related to thalamic volume. Brain Imaging Behav. 12, 20–28. doi:10.1007/s11682-016-9667-6

Bodini, B., Poirion, E., Tonietto, M., Benoit, C., Palladino, R., Maillart, E., Portera, E., Battaglini, M., Bera, G., Kuhnast, B., Louapre, C., Bottlaender, M., Stankoff, B., 2020. Individual mapping of innate immune cell activation is a candidate marker of patient-specific trajectories of disability worsening in Multiple Sclerosis. J. Nucl. Med. doi:10.2967/jnumed.119.231340

Brownlee, W., Swanton, J., Altmann, D., Ciccarelli, O., Miller, D., 2015. Earlier and more frequent diagnosis of multiple sclerosis using the McDonald criteria. J Neurol Neurosurg Psychiatry. 86, 584–585. doi:10.1136/jnnp-2014-308675

Buscarinu, M.C., Cerasoli, B., Annibali, V., Policano, C., Lionetto, L., Capi, M., Mechelli, R., Romano,

S., Fornasiero, A., Mattei, G., Piras, E., Angelini, D.F., Battistini, L., Simmaco, M., Umeton, R., Salvetti, M., Ristori, G., 2017. Altered intestinal permeability in patients with relapsing-remitting multiple sclerosis: A pilot study. Mult. Scler. 23, 442–446. doi:10.1177/1352458516652498

Calabrese, M., Pitteri, M., Farina, G., Bajrami, A., Castellaro, M., Magliozzi, R., Monaco, S., 2017. Dimethyl fumarate: A possible exit strategy from natalizumab treatment in patients with multiple sclerosis at risk for severe adverse events. J. Neurol. Neurosurg. Psychiatry 88, 1073–1078. doi:10.1136/jnnp-2017-316236

Camerota, F., Celletti, C., Di Sipio, E., De Fino, C., Simbolotti, C., Germanotta, M., Mirabella, M., Padua, L., Nociti, V., 2017. Focal muscle vibration, an effective rehabilitative approach in severe gait impairment due to multiple sclerosis. J Neurol Sci. 372, 33–39. doi:10.1016/j.jns.2016.11.025

Carotenuto, A., Costabile, T., Moccia, M., Falco, F., Scala, M.R., Russo, C. V., Saccà, F., De Rosa, A., Lanzillo, R., Brescia Morra, V., 2019. Olfactory function and cognition in relapsing–remitting and secondary-progressive multiple sclerosis. Mult. Scler. Relat. Disord. 27, 1–6. doi:10.1016/j.msard.2018.09.024

Cocozza, S., Pontillo, G., Lanzillo, R., Russo, C., Petracca, M., Di Stasi, M., Paolella, C., Vola, E.A., Criscuolo, C., Moccia, M., Lamberti, A., Monti, S., Brescia Morra, V., Elefante, A., Palma, G., Tedeschi, E., Brunetti, A., 2019. MRI features suggestive of gadolinium retention do not correlate with Expanded Disability Status Scale worsening in Multiple Sclerosis. Neuroradiology 61, 155–162. doi:10.1007/s00234-018-02150-4

Coghe, G., Fenu, G., Lorefice, L., Zucca, E., Porta, M., Pilloni, G., Corona, F., Frau, J., Giovanna Marrosu, M., Pau, M., Cocco, E., 2018. Association between brain atrophy and cognitive motor interference in multiple sclerosis. Mult. Scler. Relat. Disord. 25, 208–211.

doi:10.1016/j.msard.2018.07.045

De Biasi, S., Simone, A., Nasi, M., Bianchini, E., Ferraro, D., Vitetta, F., Gibellini, L., Pinti, M., Del Giovane, C., Sola, P., Cossarizza, A., 2016. iNKT cells in secondary Progressive Multiple sclerosis Patients Display Pro-inflammatory Profiles. Front. Immunol. 7, 555. doi:10.3389/fimmu.2016.00555

Della Corte, M., Santangelo, G., Bisecco, A., Sacco, R., Siciliano, M., D'Ambrosio, A., Docimo, R., Cuomo, T., Lavorgna, L., Bonavita, S., Tedeschi, G., Gallo, A., 2018. A simple measure of cognitive reserve is relevant for cognitive performance in MS patients. Neurol Sci. 39, 1267–1273. doi:10.1007/s10072-018-3422-2

Di Filippo, M., Proietti, S., Gaetani, L., Gubbiotti, M., Di Gregorio, M., Eusebi, P., Calabresi, P., Sarchielli, P., Giannantoni, A., 2014. Lower urinary tract symptoms and urodynamic dysfunction in clinically isolated syndromes suggestive of multiple sclerosis. Eur J Neurol. 21, 648–653. doi:10.1111/ene.12370

Di Gregorio, M., Gaetani, L., Eusebi, P., Floridi, P., Picchioni, A., Rosi, G., Mancini, A., Floridi, C., Baschieri, F., Gentili, L., Sarchielli, P., Calabresi, P., Di Filippo, M., 2018. Treatment of multiple sclerosis relapses with high-dose methylprednisolone reduces the evolution of contrast-enhancing lesions into persistent black holes. J Neurol. 265, 522–529. doi:10.1007/s00415-017-8726-2

Ferraro, D., Annovazzi, P., Moccia, M., Lanzillo, R., De Luca, G., Nociti, V., Fantozzi, R., Paolicelli, D., Ragonese, P., Gajofatto, A., Boffa, L., Cavalla, P., Lo Fermo, S., Buscarinu, M.C., Lorefice, L., Cordioli, C., Calabrese, M., Gallo, A., Pinardi, F., Tortorella, C., Di Filippo, M., Camera, V., Maniscalco, G.T., Radaelli, M., Buttari, F., Tomassini, V., Cocco, E., Gasperini, C., Solaro, C., 2020a. Characteristics and treatment of Multiple Sclerosis-related trigeminal neuralgia: An Italian multi-centre study. Mult. Scler. Relat. Disord. 37, 101461.

doi:10.1016/j.msard.2019.101461

Ferraro, D., Guicciardi, C., De Biasi, S., Pinti, M., Bedin, R., Camera, V., Vitetta, F., Nasi, M., Meletti, S., Sola, P., 2020b. Plasma neurofilaments correlate with disability in progressive multiple sclerosis patients. Acta Neurol Scand. 141, 16–21. doi:10.1111/ane.13152

Fortier, I., Raina, P., Van den Heuvel, E.R., Griffith, L.E., Craig, C., Saliba, M., Doiron, D., Stolk, R.P., Knoppers, B.M., Ferretti, V., Granda, P., Burton, P., 2017. Maelstrom Research guidelines for rigorous retrospective data harmonization. Int. J. Epidemiol. 46, 103–115. doi:10.1093/ije/dyw075

Gaetani, L., Salvadori, N., Lisetti, V., Eusebi, P., Mancini, A., Gentili, L., Borrelli, A., Portaccio, E., Sarchielli, P., Blennow, K., Zetterberg, H., Parnetti, L., Calabresi, P., Di Filippo, M., 2019. Cerebrospinal fluid neurofilament light chain tracks cognitive impairment in multiple sclerosis. J Neurol. 266, 2157–2163. doi:10.1007/s00415-019-09398-7

Gajofatto, A., Bianchi, M.R., Deotto, L., Benedetti, M.D., 2014. Are natalizumab and fingolimod analogous second-line options for the treatment of relapsing-remitting multiple sclerosis? A clinical practice observational study. Eur. Neurol. 72, 173–180. doi:10.1159/000361044

Glaser, A., Stahmann, A., Meissner, T., Flachenecker, P., Horáková, D., Zaratin, P., Brichetto, G., Pugliatti, M., Rienhoff, O., Vukusic, S., de Giacomoni, A.C.A., Battaglia, M.A.M., Brola, W., Butzkueven, H., Casey, R., Drulovic, J., Eichstädt, K., Hellwig, K., Iaffaldano, P., Ioannidou, E., Kuhle, J., Lycke, K., Magyari, M., Malbaša, T., Middleton, R., Myhr, K.M.K., Notas, K., Orologas, A., Otero-Romero, S., Pekmezovic, T., Sastre-Garriga, J., Seeldrayers, P., Soilu-Hänninen, M., Stawiarz, L., Trojano, M., Ziemssen, T., Hillert, J., Thalheim, C., 2019. Multiple Sclerosis Registries in Europe - An Updated Mapping Survey. Mult. Scler. Relat. Disord. 27, 171–178. doi:10.1016/j.msard.2018.09.032

Harris, R., Deeks, J., Altman, D., Bradburn, M., Harbord, R., Sterne, J., 2008. Metan: fixed- and

random-effects meta-analysis. Stata J. 8, 3–28.

Kalincik, T., Cutter, G., Spelman, T., Jokubaitis, V., Havrdova, E., Horakova, D., Trojano, M., Izquierdo, G., Girard, M., Duquette, P., Prat, A., Lugaresi, A., Grand'Maison, F., Grammond, P., Hupperts, R., Oreja-Guevara, C., Boz, C., Pucci, E., Bergamaschi, R., Lechner-Scott, J., Alroughani, R., Van Pesch, V., Iuliano, G., Fernandez-Bolaños, R., Ramo, C., Terzi, M., Slee, M., Spitaleri, D., Verheul, F., Cristiano, E., Sánchez-Menoyo, J.L., Fiol, M., Gray, O., Cabrera-Gomez, J.A., Barnett, M., Butzkueven, H., Grand'Maison, F., Grammond, P., Hupperts, R., Oreja-Guevara, C., Boz, C., Pucci, E., Bergamaschi, R., Lechner-Scott, J., Alroughani, R., Van Pesch, V., Iuliano, G., Fernandez-Bolaños, R., Ramo, C., Terzi, M., Slee, M., Spitaleri, D., Verheul, F., Cristiano, E., Sánchez-Menoyo, J.L., Fiol, M., Gray, O., Cabrera-Gomez, J.A., Barnett, M., Butzkueven, H., 2015. Defining reliable disability outcomes in multiple sclerosis. Brain. 138, 3287–3298. doi:10.1093/brain/awv258

Lanzillo, R., Carotenuto, A., Moccia, M., Saccà, F., Russo, C. V., Massarelli, M., De Rosa, A., Brescia Morra, V., 2017. A longitudinal real-life comparison study of natalizumab and fingolimod. Acta Neurol Scand. 136, 217–222. doi:10.1111/ane.12718

Lanzillo, R., Prosperini, L., Gasperini, C., Moccia, M., Fantozzi, R., Tortorella, C., Nociti, V., Annovazzi, P., Cavalla, P., Radaelli, M., Malucchi, S., Clerici, V.T., Boffa, L., Buttari, F., Ragonese, P., Maniscalco, G.T., Di Filippo, M., Buscarinu, M.C., Pinardi, F., Gallo, A., Coghe, G., Pesci, I., Laroni, A., Gajofatto, A., Calabrese, M., Tomassini, V., Cocco, E., Solaro, C., 2018. A multicentRE observational analysiS of PErsistenCe to Treatment in the new multiple sclerosis era: the RESPECT study. J Neurol. 265, 1174–1183. doi:10.1007/s00415-018-8831-x

Lipp, I., Parker, G.D., Tallantyre, E.C., Goodall, A., Grama, S., Patitucci, E., Heveron, P., Tomassini, V., Jones, D.K., 2020. Tractography in the presence of multiple sclerosis lesions. Neuroimage 209, 116471. doi:10.1016/j.neuroimage.2019.116471

Lorefice, L., Fenu, G., Sardu, C., Frau, J., Coghe, G., Costa, G., Schirru, L., Secci, M.A., Sechi, V., Barracciu, M.A., Marrosu, M.G., Cocco, E., 2019. Multiple sclerosis and HLA genotypes: A possible influence on brain atrophy. Mult. Scler. 25, 23–30. doi:10.1177/1352458517739989

Mallucci, G., Annovazzi, P., Miante, S., Torri-Clerici, V., Matta, M., La Gioia, S., Cavarretta, R., Mantero, V., Costantini, G., D'Ambrosio, V., Zaffaroni, M., Ghezzi, A., Perini, P., Rossi, S., Bertolotto, A., Rottoli, M.R., Rovaris, M., Balgera, R., Cavalla, P., Montomoli, C., Bergamaschi, R., 2018. Two-year real-life efficacy, tolerability and safety of dimethyl fumarate in an Italian multicentre study. J Neurol. 265, 1850–1859. doi:10.1007/s00415-018-8916-6

Malucchi, S., Capobianco, M., Lo Re, M., Malentacchi, M., di Sapio, A., Matta, M., Sperli, F., Bertolotto, A., 2017. High-Risk PML Patients Switching from Natalizumab to Alemtuzumab: an Observational Study. Neurol. Ther. 6, 145–152. doi:10.1007/s40120-016-0058-0

Manni, A., Iaffaldano, A., Lucisano, G., D'Onghia, M., Mezzapesa, D.M., Felica, V., Iaffaldano, P., Trojano, M., Paolicelli, D., 2019. Lymphocyte count and body mass index as biomarkers of early treatment response in a multiple sclerosis dimethyl fumarate-treated cohort. Front. Immunol. 10, 1343. doi:10.3389/fimmu.2019.01343

Manouchehrinia, A., Westerlind, H., Kingwell, E., Zhu, F., Carruthers, R., Ramanujam, R., Ban, M., Glaser, A., Sawcer, S., Tremlett, H., Hiller, J., 2017. Age Related Multiple Sclerosis Severity Score: Disability ranked by age. Mult Scler 23, 1938–1946. doi:10.1177/1352458516688

Marastoni, D., Buriani, A., Pisani, A.I., Crescenzo, F., Zuco, C., Fortinguerra, S., Sorrenti, V., Marenda, B., Romualdi, C., Magliozzi, R., Monaco, S., Calabrese, M., 2019. Increased NK Cell Count in Multiple Sclerosis Patients Treated With Dimethyl Fumarate: A 2-Year Longitudinal Study. Front. Immunol. 10, 1666. doi:10.3389/fimmu.2019.01666

Megna, R., Alfano, B., Lanzillo, R., Costabile, T., Comerci, M., Vacca, G., Carotenuto, A., Moccia, M., Servillo, G., Prinster, A., Brescia Morra, V., Quarantelli, M., 2019. Brain tissue volumes and

relaxation rates in multiple sclerosis: implications for cognitive impairment. J Neurol. 266, 361–368. doi:10.1007/s00415-018-9139-6

Middleton, R.M., Rodgers, W.J., Chataway, J., Schmierer, K., Rog, D., Galea, I., Akbari, A., Tuite-Dalton, K., Lockhart-Jones, H., Griffiths, D., Noble, D.G., Jones, K.H., Al-Din, A., Craner, M., Evangelou, N., Harman, P., Harrower, T., Hobart, J., Husseyin, H., Kasti, M., Kipps, C., McDonnell, G., Owen, C., Pearson, O., Rashid, W., Wilson, H., Ford, D. V., 2018. Validating the portal population of the United Kingdom Multiple Sclerosis Register. Mult. Scler. Relat. Disord. 24, 3–10. doi:10.1016/j.msard.2018.05.015

Moccia, M., Capacchione, A., Lanzillo, R., Carbone, F., Micillo, T., Perna, F., De Rosa, A., Carotenuto, A., Albero, R., Matarese, G., Palladino, R., Brescia Morra, V., 2019. Coenzyme Q10 supplementation reduces peripheral oxidative stress and inflammation in Interferon-Beta1a treated multiple sclerosis. Ther. Adv. Neurol. Disord. 12, 1–12. doi:10.1177/1756286418819074

Moccia, M., Lanzillo, R., Palladino, R., Chang, K.K.C.M.K.C.-M., Costabile, T., Russo, C., De Rosa, A., Carotenuto, A., Saccà, F., Maniscalco, G.T.G.T., Brescia Morra, V., 2016. Cognitive impairment at diagnosis predicts 10-year multiple sclerosis progression. Mult. Scler. 22, 659–667. doi:10.1177/1352458515599075

Paolicelli, D., Direnzo, V., Manni, A., D'Onghia, M., Tortorella, C., Zoccolella, S., Di Lecce, V., Iaffaldano, A., Trojano, M., 2016. Long-Term Data of Efficacy, Safety, and Tolerability in a Real-Life Setting of THC/CBD Oromucosal Spray-Treated Multiple Sclerosis Patients. J. Clin. Pharmacol. 56, 845–851. doi:10.1002/jcph.670

Pellegrino, L., Coscia, M., Muller, M., Solaro, C., Casadio, M., 2018. Evaluating upper limb impairments in multiple sclerosis by exposure to different mechanical environments. Sci. Rep. 8, 1–14. doi:10.1038/s41598-018-20343-y

Petruzzo, M., Palladino, R., Nardone, A., Nozzolillo, A., Servillo, G., Orlando, V., De Angelis, M., Lanzillo, R., Brescia Morra, V., Moccia, M., 2020. The impact of diagnostic criteria and treatments on the 20-year costs for treating relapsing-remitting multiple sclerosis. Mult. Scler. Relat. Disord. 38, 101514. doi:10.1016/j.msard.2019.101514

Pugliatti, M., Eskic, D., Mikolcić, T., Pitschnau-Michel, D., Myhr, K.M., Sastre-Garriga, J., Otero, S., Wieczynska, L., Torje, C., Holloway, E., Rienhoff, O., Friede, T., Buckow, K., Ellenberger, D., Hillert, J., Glaser, A., Flachenecker, P., Fuge, J., Schyns-Liharska, T., Kasilingam, E., Moretti, A., Thalheim, C., 2012. Assess, compare and enhance the status of Persons with Multiple Sclerosis (MS) in Europe: A European Register for MS. Acta Neurol Scand. 126, 24–30. doi:10.1111/ane.12024

Realmuto, S., Dodich, A., Meli, R., Canessa, N., Ragonese, P., Salemi, G., Cerami, C., 2019. Moral Cognition and Multiple Sclerosis: A Neuropsychological Study. Arch. Clin. Neuropsychol. 34, 319–326. doi:10.1093/arclin/acy047

Salemi, G., Vazzoler, G., Ragonese, P., Bianchi, A., Cosentino, G., Croce, G., Gangitano, M., Portera, E., Realmuto, S., Fierro, B., Brighina, F., 2019. Application of tRNS to improve multiple sclerosis fatigue: a pilot, single-blind, sham-controlled study. J. Neural Transm. 126, 795–799. doi:10.1007/s00702-019-02006-y

Salter, A., Stahmann, A., Ellenberger, D., Fneish, F., Rodgers, W.J., Middleton, R., Nicholas, R., Marrie, R.A., 2020. Data harmonization for collaborative research among MS registries: A case study in employment. Mult. Scler. doi:10.1177/1352458520910499

Scalfari, A., Romualdi, C., Nicholas, R.S., Mattoscio, M., Magliozzi, R., Morra, A., Monaco, S., Muraro, P.A., Calabrese, M., 2018. The cortical damage, early relapses, and onset of the progressive phase in multiple sclerosis. Neurology 90, e2107–e2118. doi:10.1212/WNL.0000000000005685

Schwehr, N.A., Kuntz, K.M., Butler, M., Enns, E.A., Shippee, N.D., Kingwell, E., Tremlett, H., Carpenter, A.F., 2018. Age-related decreases in relapses among adults with relapsing-onset multiple sclerosis. Mult. Scler. 24, 642–652. doi:10.1177/1352458519866613

Signori, A., Izquierdo, G., Lugaresi, A., Hupperts, R., Grand'Maison, F., Sola, P., Horakova, D., Havrdova, E., Prat, A., Girard, M., Duquette, P., Boz, C., Grammond, P., Terzi, M., Singhal, B., Alroughani, R., Petersen, T., Ramo, C., Oreja-Guevara, C., Spitaleri, D., Shaygannejad, V., Butzkueven, H., Kalincik, T., Jokubaitis, V., Slee, M., Fernandez Bolaños, R., Sanchez-Menoyo, J.L., Pucci, E., Granella, F., Lechner-Scott, J., Iuliano, G., Hughes, S., Bergamaschi, R., Taylor, B., Verheul, F., Edite Rio, M., Amato, M.P., Sajedi, S.A., Majdinasab, N., Van Pesch, V., Sormani, M.P., Trojano, M., 2018. Long-term disability trajectories in primary progressive MS patients: A latent class growth analysis. Mult. Scler. 24, 642–652. doi:10.1177/1352458517703800

Sola, P., Mandrioli, J., Simone, A., Ferraro, D., Bedin, R., Annecca, R., Venneri, M., Nichelli, P., Merelli, E., 2010. Primary progressive versus relapsing-onset multiple sclerosis: presence and prognostic value of cerebrospinal fluid oligoclonal IgM. Mult. Scler. 17, 303–311. doi:DOI: 10.1177/135245851038699

Thompson, A.J., Baranzini, S.E., Geurts, J., Hemmer, B., Ciccarelli, O., 2018. Multiple sclerosis. Lancet 391, 1622–1636. doi:10.1016/S0140-6736(18)30481-1

Trojano, M., Tintore, M., Montalban, X., Hillert, J., Kalincik, T., Iaffaldano, P., Spelman, T., Sormani, M., Butzkueven, H., 2017. Treatment decisions in multiple sclerosis - insights from real-world observational studies. Nat Rev Neurol 13, 105–118. doi:10.1038/nrneurol.2016.188

Trojano, Maria, Bergamaschi, Roberto, Amato, Maria Pia, Comi, Giancarlo, Ghezzi, A., Lepore, V., Marrosu, M.G., Mosconi, P., Patti, Francesco, Ponzio, M., Zaratin, P., Battaglia, M.A., Acquistapace, D., Aguglia, U., Amato, M. P., Annunziata, P., Ardito, B., Avolio, C., Balgera, R., Bandini, F., Banfi, P., Barone, P., Bellantonio, P., Bergamaschi, R., Bertolotto, A., Bertora, P.,

Bombardi, R., Bosco Zimatore, G., Bossio, R.B., Bramanti, P., Brescia Morra, V., Brioschi, A.M., Bruzzone, M., Buccafusca, M., Busillo, V., Caneve, G., Caniatti, L.M., Capone, L., Capone, F., Cappellani, A., Cargnelutti, D., Cavaletti, G., Cavalla, P., Celani, M.G., Centonze, D., Chiveri, L., Clerici, R., Clerico, M., Cocco, E., Comi, G., Comi, C., Coniglio, M.G., Cordera, S., Corea, F., Cortese, A., Costantino, G., Cottone, S., Crociani, P., D'Andrea, F., Danni, M.C., De Luca, G., de Pascalis, D., De Robertis, F., De Stefano, N., Di Battista, G., Di Napoli, M., Falcini, M., Fausto, F., Ferrò, M.T., Florio, C., Fortunato, M., Frittelli, C., Galgani, S., Gallo, P., Gatto, M., Gazzola, P., Geda, C., Giordano, A., Granella, F., Grasso, M.G., Grimaldi, L.M.E., Imperiale, D., Lo Russo, L., Logullo, F.O., Lugaresi, A., Lus, G., Maccarrone, G., Maimone, D., Malagù, S., Marconi, R., Maritato, P., Massacesi, L., Mazzoni, M., Meucci, G., Mirabella, M., Montepietra, S., Nasuelli, D., Neri, W., Orefice, G., Parodi, S., Pasquali, L., Passarella, B., Patti, F., Peresson, M., Perla, F., Pesci, I., Piantadosi, C., Piras, M.L., Pizio, N.R., Pozzilli, C., Protti, A., Pugliatti, M., Quatrale, R., Ragno, M., Ragno, M., Rezzonico, M., Ribizzi, G., Riva, M., Ronzoni, M., Rosso, M.G., Rottoli, M., Rovaris, M., Salemi, G., Salvetti, M., Santangelo, M., Santangelo, G., Santuccio, G., Santuccio, G., Sarchielli, P., Scarpini, E., Sechi, G.P., Severi, S., Sinisi, L., Sola, P., Spitaleri, D., Tassinari, T., Tedeschi, G., Tonietti, S., Torri Clerici, V., Totaro, R., Traccis, S., Trojano, M., Turla, M., Uccelli, A., Ulivelli, M., Valentino, P., Valeriani, M., Venturi, S., Vianello, M., Zaffaroni, M., 2019. The Italian multiple sclerosis register. Neurol Sci. 40, 155–165. doi:10.1007/s10072-018-3610-0

Tur, C., Moccia, M., Barkhof, F., Chataway, J., Sastre-Garriga, J., Thompson, A.J., Ciccarelli, O., 2018. Assessing treatment outcomes in multiple sclerosis trials and in the clinical setting. Nat. Rev. Neurol. 14, 75–93. doi:10.1038/nrneurol.2017.171

Vercellino, M., Romagnolo, A., Mattioda, A., Masera, S., Piacentino, C., Merola, A., Chiò, A., Mutani, R., Cavalla, P., 2009. Multiple sclerosis relapses: A multivariable analysis of residual disability determinants. Acta Neurol Scand. 119, 126–130. doi:10.1111/j.1600-0404.2008.01076.x

Wan, X., Wang, W., Liu, J., Tong, T., 2014. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med. Res. Methodol. 14, 135. doi:10.1186/1471-2288-14-135

Zoccolella, S., Tortorella, C., Iaffaldano, P., Direnzo, V., D'Onghia, M., Paolicelli, D., Livrea, P., Trojano, M., 2012. Elevated plasma homocysteine levels in patients with multiple sclerosis are associated with male gender. J Neurol. 259, 2105–2110. doi:10.1007/s00415-012-6464-z

Figure 1. Modular sets of outcome measures

Figure shows modular sets of outcome measures. Basic modules for diagnosis and for follow-up consisted of outcome measures recorded by all participating centres at the time of diagnosis, and, then, at least annually, during follow-up. The detailed module included outcome measures recorded in most participating centres in the presence of specific clinical indications. The optional extension module included outcome measures recorded in some participating centres in the presence of specific clinical indications.

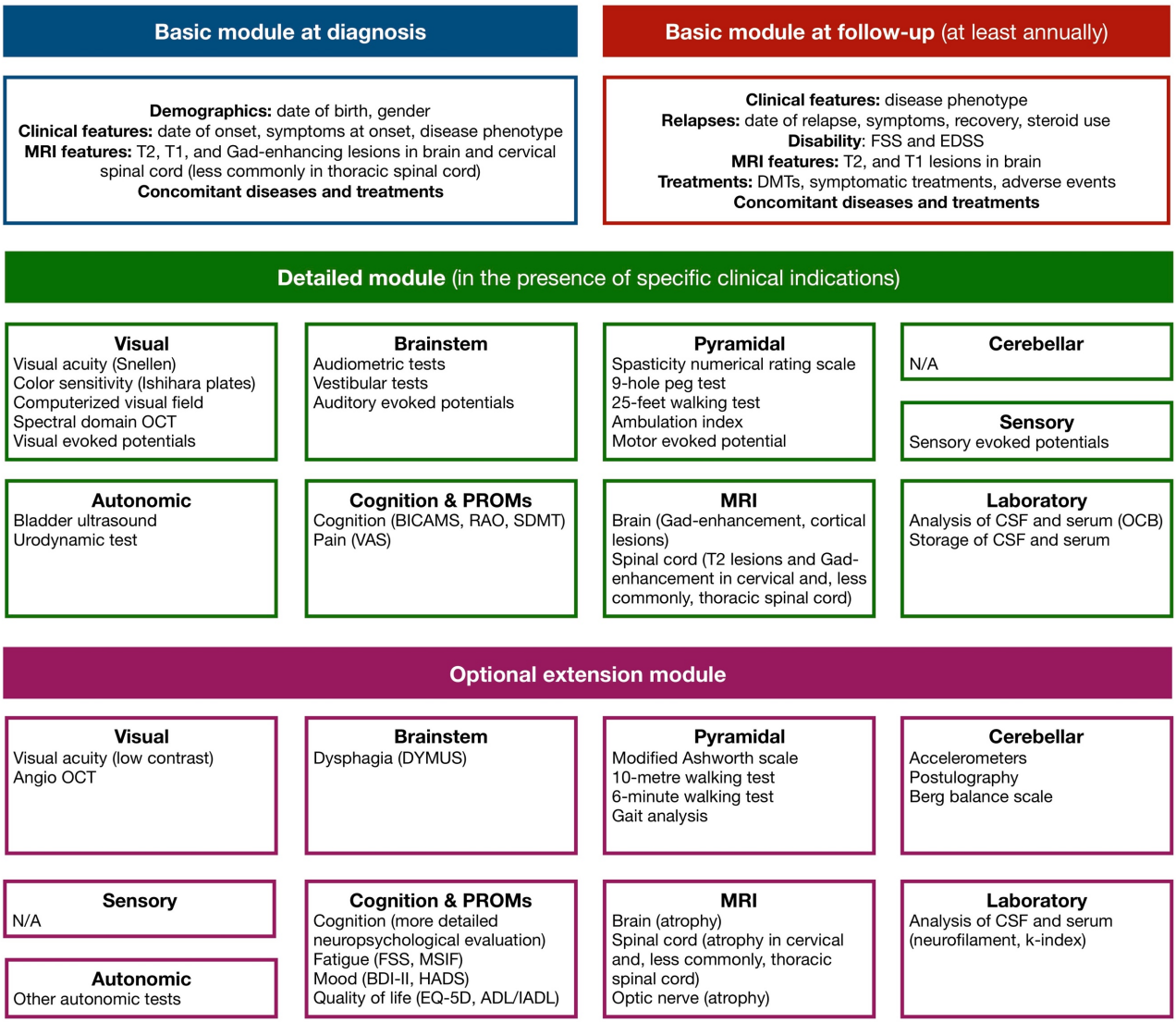

Figure 2. Heterogeneity of demographic features.

Forest plots show heterogeneity of age at onset (**a**), and female percent (**b**) between studies including RRMS patients, estimated using the DerSimonian and Laird random-effects method. Results are presented as effect size, 95% confidence intervals (95%CI), and I^2.

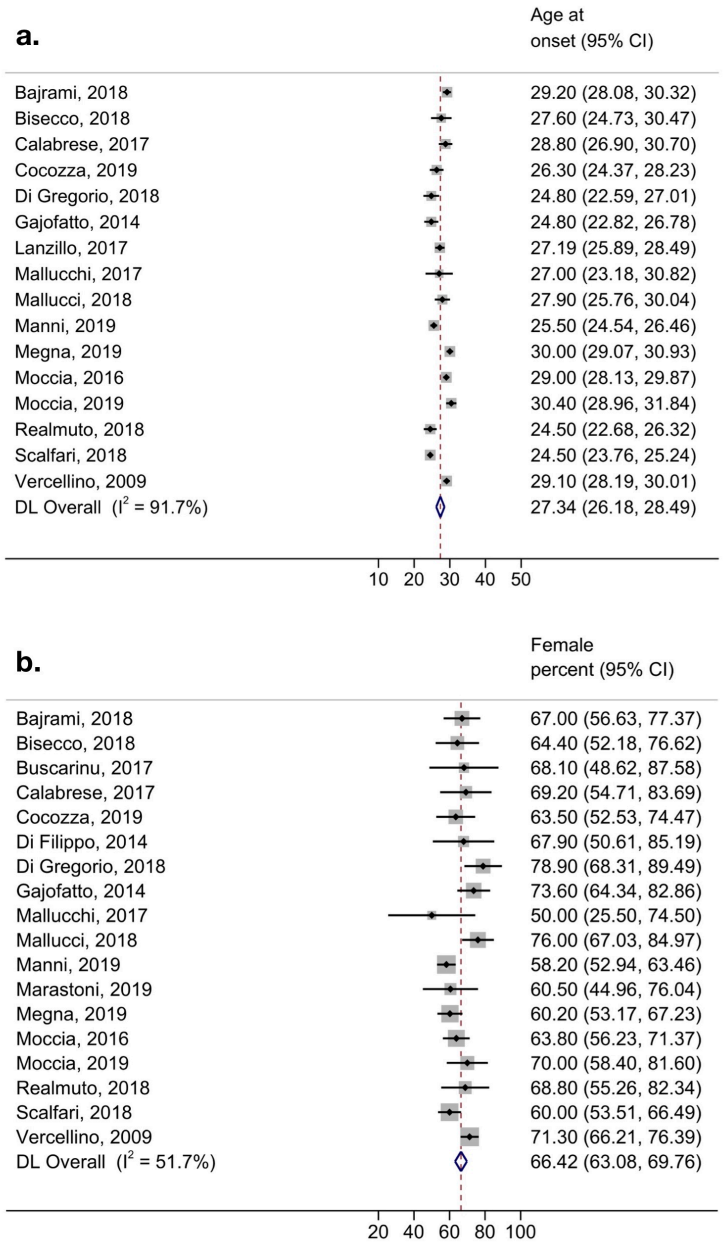

Figure 3. Heterogeneity of baseline EDSS.

Forest plots show heterogeneity of baseline EDSS between studies, in subgroups of different age ranges, estimated using the DerSimonian and Laird random-effects method. Results are presented as effect size, 95% confidence intervals (95%CI), and I^2.

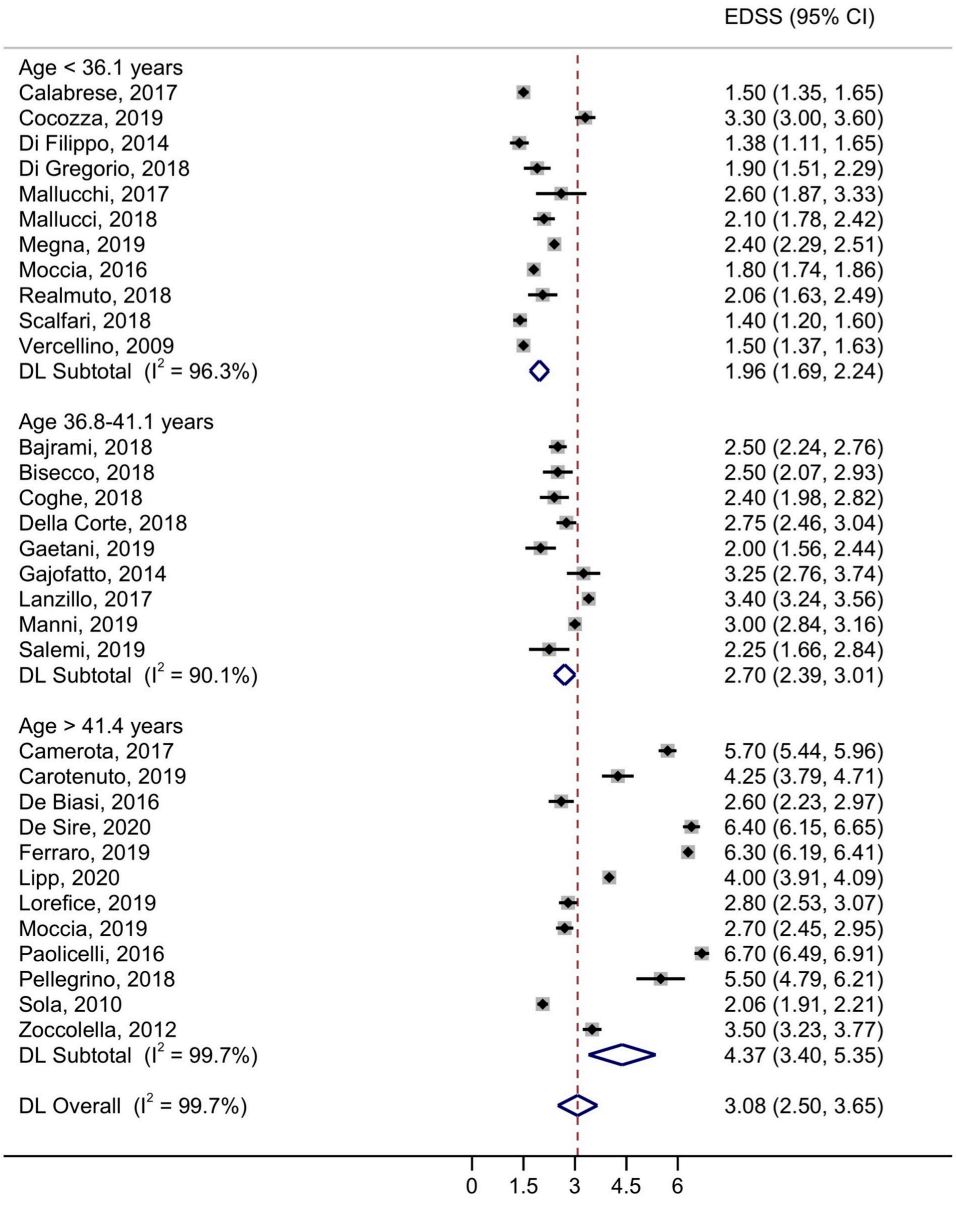

Figure 4. Heterogeneity of longitudinal outcomes.

Forest plots show heterogeneity of yearly rates in relapse occurrence (**a**), and EDSS progression between longitudinal studies (**b**), in subgroups of different study durations, estimated using the DerSimonian and Laird random-effects method. Results are presented as effect size, 95% confidence intervals (95%CI), and I^2.

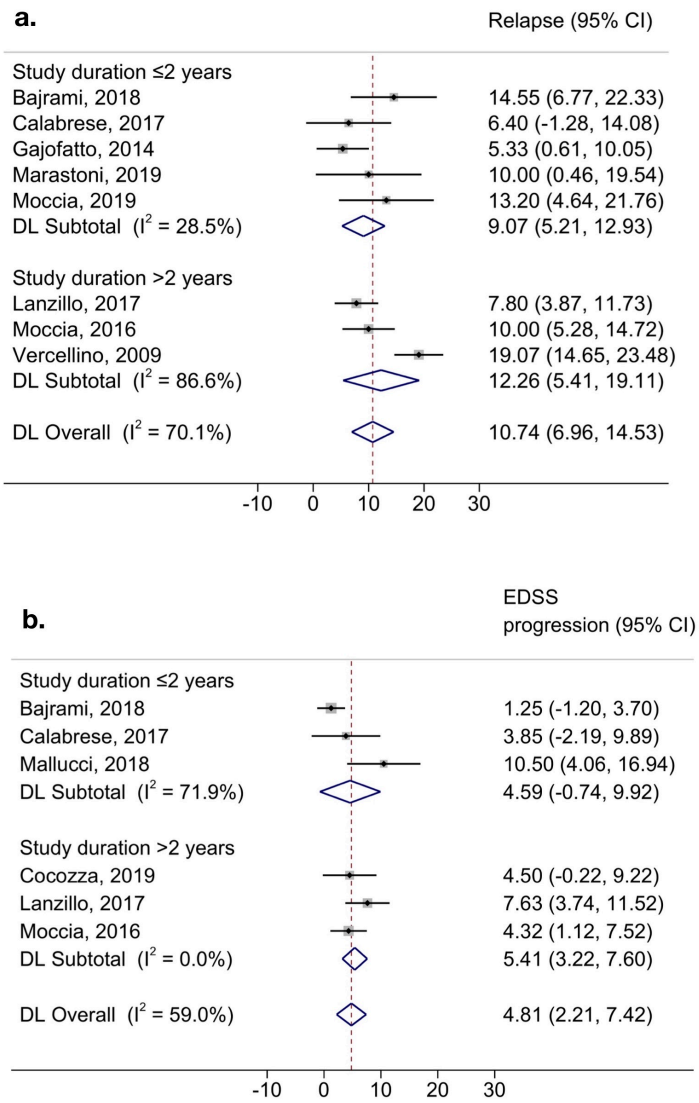

Figure 5. Maelstrom Research guidelines with MS-specific recommendations.

Figure shows MS-specific recommendations for rigorous retrospective data harmonization, in line with the Maelstrom Research guidelines.

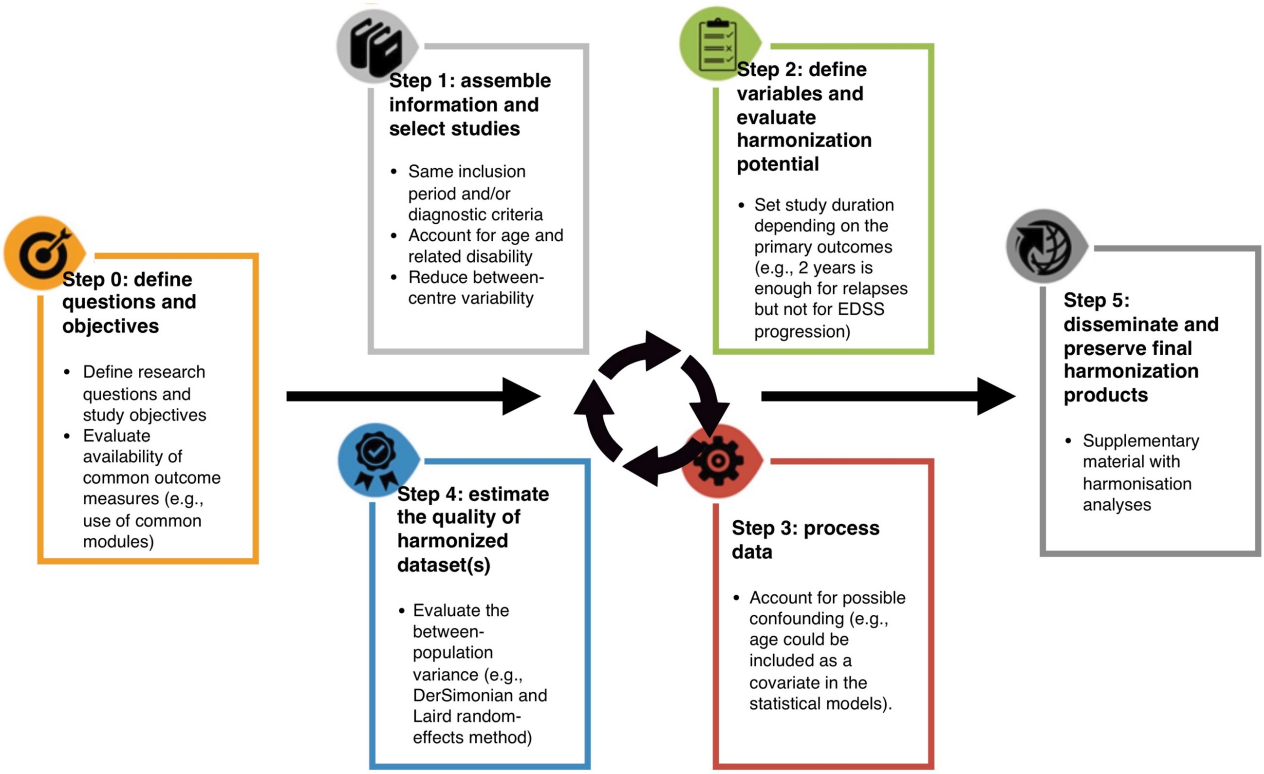

Table 1. Studies included for heterogeneity analyses.

Table shows studies included for heterogeneity analyses, with details on study design, demographics and clinical features.

Reference	Disease subtype	Sample Size *(n)*	Study Duration *(years)*	Age at onset *(years)*	Female *(%)*	EDSS	T25FWT *(seconds)*	9HPT *(seconds)*	Relapses *(%/year)*	EDSS progression *(%/year)*	Brain volume *(mL)*	GM volume *(mL)*	SDMT	PASAT3	BDI	FSS	MFIS
Bajrami et al. J Neurol 2018	RRMS	79	2.0	29.2±5.1	67.0	2.5±1.1			14.6	1.3							
Bisecco et al. Brain Imaging Behav 2018	RRMS	59		27.6±11.2	64.4	2.5±1.7					1519.0±80.0	789.0±62.0					
Buscarinu et al. Mult Scler 2017	RRMS	22			68.1												
Calabrese et al. JNNP 2017	RRMS	39	2.0	28.8±6.0	69.2	1.5±0.4			6.4	3.9							
Camerota et al. J Neurol Sci 2017	PMS	14			42.8	5.7±0.5	19.7±13.5								5.5±3.4		38.5±18.4
Carotenuto et al. Mult Scler Relat Disord 2019	Mixed	55		30.2±13.1	69.0	4.2±1.7	13.3±10	29.9±16.4					38.9±16.0	26.8±20.7	14.4±10.3		36.9±18.7
Cocozza et al. Neuroradiology 2019	RRMS	74	3.6	26.3±8.4	63.5	3.3±1.3				4.5		763.2±61.2					
Coghe et al. Mult Scler Relat Disord 2018	Mixed	48			68.7	2.4±1.5											
De Biasi et al. Front Immunol 2016	Mixed	165		33.8±9.5	74.0	2.6±2.4											
De Sire et al. NeuroRehabiltation 2019	Mixed	10		35.0±7.4	60.0	6.4±0.4											
Della Corte et al. Neurol Sci 2018	Mixed	147		27.7±10.1	72.7	2.7±1.7							35.3±14.4	34.8±13.7		32.4±16.0	
Di Filippo et al. Eur J Neurol 2014	RRMS	28			67.9	1.3±0.7											

Di Gregorio et al. J Neurol 2018	RRMS	57	2.1	24.8±8.5	78.9	1.9±1.5												
Ferraro et al. Acta Neurol Scand 2019	PMS	70	4.6	39.2±2.9	70.0	6.3±0.4												
Gaetani et al. J Neurol 2019	Mixed	28		36.5±10.8	60.0	2.0±1.2									42.6±12	43±14.1		
Gajofatto et al. Eur. Neurol 2014	RRMS	87	1.5	24.8±9.4	73.6	3.2±2.3			5.3									
Lanzillo et al. Acta Neurol Scand 2017	RRMS	179	3.0	27.1±8.9		3.4±1.0			7.8	7.6								
Lipp et al. Neuroimage 2020	Mixed	131		32.1±7.5	64.0	4.0±0.5	8.5±9.8	25.4±11.7			1173.5±115.9	594.4±63.2		39.9±14.0	12.5±10.4		39.6±20.7	
Lorefice et al. Mult Scler 2019	Mixed	240		29±7.4	72.0	2.8±2.1					1444.8±85.0	763.1±63.7						
Mallucchi et al. Neurol Ther 2017	RRMS	16	1.5	27±7.8	50.0	2.6±1.5												
Mallucci et al. J Neurol 2018	RRMS	87	2.0	27.9±10.2	76.0	2.1±1.5				10.5								
Manni et al. Front Immunol 2019	RRMS	338		25.5±9	58.2	3.0±1.5												
Marastoni et al. Front Immunol 2019	RRMS	38	2.0		60.5				10.0									
Megna et al. J Neurol 2019	RRMS	186		30.0±6.5	60.2	2.4±0.8							48.4±14.6	35.4±13.3				
Moccia et al. Mult Scler 2016	RRMS	155	10.0	29.0±5.5	63.8	1.8±0.4			10.0	4.3			30.6±14.6	28.5±10.6				
Moccia et al. Ther Adv Neurol Disord 2019	RRMS	60	0.5	30.4±5.7	70.0	2.7±1.0			13.2						14.1±10.2	27.5±10.2		
Paolicelli et al. J Clin Pharmacol 2016	Mixed	102		29.6±9.2	49.0	6.7±1.1	28.1±17.2											
Pellegrino et al. Sci Rep 2018	Mixed	11			81.0	5.5±1.2		47.8±24.2								47.2±11.7		
Realmuto et al.	RRMS	45		24.5±6.2	68.8	2.0±1.4							47.4±10.7			61.0±35.1		

Arch Clin Neuropsychol 2018											
Salemi et al.	Mixed	9	0.2	32.1±7.7	66.0	2.2±0.9	8.0±1.6			39±13.7	42.3±13.1
J Neural Transm 2019											
Scalfari et al.	RRMS	219	7.9	24.5±5.6	60.0	1.4±1.5					
Neurology 2018											
Sola et al.	Mixed	149		44.4±6.7	60.0	2.0±0.9					
Mult Scler 2010											
Vercellino et al.	RRMS	304	3.0	29.1±8.1	71.3	1.5±1.1		19.1			
Acta Neurol Scand 2009											
Zoccolella et al.	Mixed	217		32.1±13.5	65.4	3.5±2.0					
J Neurol 2012											