Title: Validating the Portal Population of the United Kingdom Multiple Sclerosis Register

Authors: Middleton RM[1], Rodgers WJ[1], Chataway J[2], Schmierer K[3], Rog D[4], Galea I[5], Akbari A[1], Tuite-Dalton K[1], Lockhart-Jones H[1], Griffiths D[1], Noble DG[1], Jones KH[1], Al-Din A[6], Craner M[7], Evangelou N[8], Harman P[9], Harrower T[10], Hobart J[11], Husseyin H[12], Kasti M[13] , Kipps C[14], McDonnell G[15] , Owen C[16], Pearson O[17] , Rashid W[18], Wilson H[19], Ford DV[1]

Corresponding: Author r.m.middleton@swansea.ac.uk

Institutions:

[1] Swansea University Medical School, Swansea, United Kingdom
[2] University College London, London, United Kingdom
[3] Queen Mary University of London, Blizard Institute, London, United Kingdom
[4] Salford Royal Hospital, Manchester, United Kingdom
[5] University of Southampton, Clinical Neurosciences, Southampton, United Kingdom
[6] National Health Service, Dept of Neurology, North Yorkshire, United Kingdom
[7] University of Oxford, Nuffield Department of Clinical Neurosciences, Oxford, United Kingdom
[8] University of Nottingham, Division of Clinical Neurology, Nottingham, United Kingdom
[9] Southend University Hospital NHS Foundation Trust, Southend, United Kingdom
[10] Exeter Foundation Trust, Department of Neurology Royal Devon, Exeter, United Kingdom
[11] Peninsula College of Medicine and Dentistry, Plymouth, United Kingdom
[12] Luton and Dunstable NHS Foundation Trust, Luton, United Kingdom
[13] Basildon and Thurrock Hospitals NHS Foundation Trust, Basildon United Kingdom
[14] University of Southampton, Faculty of Medicine, Southampton, United Kingdom
[15] Royal Group of Hospitals, Department of Neurology, Belfast, United Kingdom
[16] Shrewsbury and Telford Hospital NHS Trust, Shrewsbury, United Kingdom
[17] Abertawe Bro Morgannwg University Local Health Board, Swansea, United Kingdom
[18] Brighton and Sussex University Hospitals NHS Trust, Department of Neurology, Brighton, United Kingdom
[19] King's College London, London, United Kingdom

Declarations of interest: none

ABSTRACT

The UK Multiple Sclerosis Register (UKMSR) is a large cohort study designed to capture 'real world' information about living with multiple sclerosis (MS) in the UK from diverse sources. The primary source of data is directly from people with Multiple Sclerosis (pwMS) captured by longitudinal questionnaires via an internet portal. This population's diagnosis of MS is self-reported and therefore unverified. The second data source is clinical data which is captured from MS Specialist Treatment centres across the UK. This includes a clinically confirmed diagnosis of MS (by Macdonald criteria) for consented patients.

A proportion of the internet population have also been consented at their hospital making comparisons possible. *This dataset is called the 'linked dataset'.* The purpose of this paper is to examine the characteristics of the three datasets: the self-reported portal data, clinical data and linked data, in order to assess the validity of the self-reported portal data.

The internet (n=11,021) and clinical (n=3,003) populations were studied for key shared characteristics. We found them to be closely matched for mean age at diagnosis (clinical=37.39, portal=39.28) and gender ratio (female %, portal=73.1, clinical=75.2). The Two Sample Kolmogorov-Smirnov test was for the continuous variables to examine is they were drawn from the same distribution. The null hypothesis was rejected only for age at diagnosis ($D = 0.078$, $p < 0.01$). The populations therefore, were drawn from different distributions, as there are more patients with relapsing disease in the clinical cohort. In all other analyses performed, the populations were shown to be drawn from the same distribution.

Our analysis has shown that the UKMSR portal population is highly analogous to the entirely clinical (validated) population. This supports the validity of the self-reported diagnosis and therefore that the portal population can be utilised as a viable and valid cohort of people with Multiple Sclerosis for study.

Keywords: Multiple Sclerosis, data linkage, longitudinal, research register, validation, PROMs

1. Introduction

Multiple sclerosis (MS) is an inflammatory demyelinating and degenerative disease of the central nervous system (CNS), and the most common non-traumatic cause of disability in young adults worldwide (1). The dominant phenotype is characterised by relapses (attacks) and remissions - relapsing MS (RRMS). This evolves at 10-15 years into secondary progressive MS (SPMS) in the majority of those affected. About 15% of people with MS (PwMS) develop progressive neurological dysfunction from onset – primary progressive MS (PPMS).(2)

The disease affects about 120,000 people in the United Kingdom (UK) with an incidence that appears to be increasing by approximately 2.5% per annum (4). There is a notably significant societal and global burden of £3bn/year (3)

In order to comprehensively map the prevalence and characteristics of MS across the UK, the MS Society commissioned Swansea University Medical School, home to the Secure Anonymous Information Linkage (SAIL) Databank (4), to develop the UK Multiple Sclerosis Register (UKMSR). The SAIL databank is a safe haven for billions of person-based records combined with a complete data linkage and an analysis toolset. The knowledge and resources used to operate and maintain this SAIL were instrumental in the development of the UKMSR

Data to be captured were established during a setup and pilot phase of 36 months following a review of data captured by a number of European Registries (5) and reference to UK Neurologists, research academics and PwMS. The 'real world' data for the UKMSR would be captured from three primary sources:

1) Data provided by PwMS via the internet (portal data).
2) Data from NHS hospitals (clinical data).
3) Data mined from general practice and inpatient hospital records (routine data).

The Register has a mandated minimum clinical dataset [Appendix 2] which was developed by its Clinical Advisory Group (CAG) of UK neurologists. This pragmatic dataset, which is collected annually, following patient consent. Contains both demographic and clinical variables. Clinical data are then linked with portal data. Linkage occurs either when the participant enters their Study ID, or through deterministic and probabilistic methodologies (6). This route was chosen as there are known, notable limitations in finding MS patients in inpatient, outpatient and General Practice data. Secondary care routinely capture patient events by ICD10 codes (G35-G37) as "Demyelinating disease of the Central Nervous System" (7) and general practitioner systems a READ code as "Multiple Sclerosis"(F20) (8). With no finer grained detail as to disease type available, this makes the minimum dataset capture of MS Specialist treatment centre data an essential part of the validation process.

Other studies using solely participant supplied data such as the North American Research Committee on Multiple Sclerosis (NARCOMS) were successful in validating their populations by carrying out expert review on a sample of their population. A selection of participants were consented, their medical records requested and treating physicians interviewed. The observed eligible population for this validation exercise was 142 participants out of 30,691, this further reducing to 52 consenting to participate.(9). Despite some caveats, NARCOMS reported the diagnostic accuracy of their population as being 98.7±1.3% in 2006. Since UKMSR portal participant's self-reported data can be validated with their clinically collected data, there was no need to contact clinicians or review patient notes as clinical data is transmitted from source.

2. Methods

2.1 Study design and participants

The UKMSR has two main sources of data: Patient Reported Outcome Measures (PROMs) and demographics supplied via the web portal and secondly consented patients 'clinical data' from MS Specialist Treatment centres across the UK. Participants on both platforms are encouraged to join the other where possible. That is PwMS who give informed consent for the transmission of their clinical data are encouraged to also sign up to the portal and portal patients who are treated at an NHS partner site are encouraged to give their consent.

This ultimately creates three data sets for analysis, the portal, the clinical and the linked population that exist in both primary sources [Figure 1].

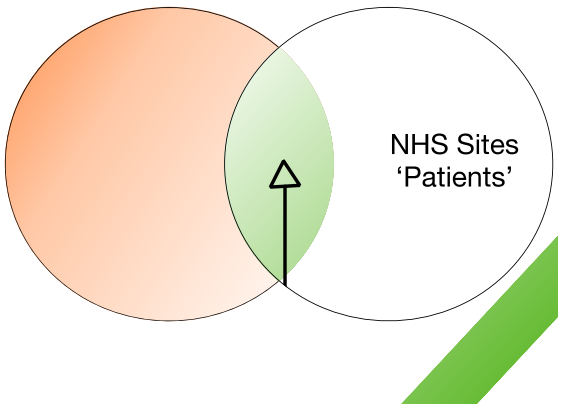

Figure 1- Venn diagram showing the linked population as an intersection of the internet participants with consented patients from a clinical site.

The purpose of this paper is to examine the characteristics of these datasets in order to assess the validity of the self-reported portal data

2.11 Portal participants

Portal participants must be aged 18 years or older and resident within the UK, they agree to a terms of service (10) stating the Register's responsibility for the use and storage of their data. PROMs are then presented in order to obtain demographics and other MS specific data. Participants then receive three monthly reminders requesting they return to complete new instruments. Initial PROMs include the Hospital Anxiety and Depression Scale (HADS) (11), EurQol Quality of Life (EQ-5D-3L) (12) and the Multiple Sclerosis Impact Scale (MSIS29v2) (13) Up to five years of longitudinal data for some participants have now been captured.

2.12 Clinical participants

Eligibility criteria for consent via an NHS Clinic requires that patients are aged over 18, resident in the UK and have a diagnosis of MS from a neurologist according to the McDonald Criteria (14). The UKMSR has been approved as a study by the South West Central Bristol Research Ethics Council, initially under registration code 11/SW/0160. Approval was renewed after five years under registration code 16/SW/0194. Consent is usually followed by collection of the minimum dataset [appendix 2]. This dataset comprises demography, disease history and course, symptoms, relapses and progression (if any), current and previous disease modifying therapies and two measures of function; a timed walk and clinician supplied EDSS score.

Participating clinical sites [Appendix 1] are self-selecting, based on an application to the UKMSR directly, patient consent is taken, and data capture begins. Data are transmitted to the UKMSR using encrypted secure transfer methods, only electronic data are accepted, and uploads are carried out by sites monthly.

2.13 Linked participants

Consented patients receive a unique study ID, which they are asked to enter onto the portal to link their data. Although probabilistic matching could be used to link patients between their clinical and portal data, only exact matches using the individually assigned study ID were used for this paper. Probabilistic matching relies on 'fuzzy' logic that build up a degree of confidence in the match, giving a probability or likelihood of match. This requires more than one variable between datasets, normally name, postcode, gender and date of birth. A score is assigned to each match, ranking the chance that it is a true match. Matches above a certain level of confidence, i.e. >95 % can then be used for the linkage. Deterministic matching, which requires an exact match between fields is the most exact method and was used for this study.

2.2 Data

All data were stored on a Microsoft SQL Server 2014 Database and Structured Query Language (SQL) queries were made to identify data for initial analysis, data cleansing and aggregation,

prior to statistical analysis. Analysis was conducted using the R statistical computing programming language (15) in the RStudio environment.

Simple descriptive statistics were used to compare the portal and clinical datasets: mean (standard deviation) for continuous data and frequency tables for categorical data on key demographic markers. A Two-Sample Kolmogorov-Smirnov (K-S) test was then implemented using age at diagnosis and current age. This is a non-parametric statistical test which determines if two different continuous variables are from the same distribution.

2.21 Portal Data

In 2012, Jones et al described the portal population (16) as having unique 7,279 participants by March 2017 this number has almost doubled to 14,720. After removing deceased, left study, obvious input errors (such as diagnosis in 1927) or those with incomplete eligible data, the number of records was reduced to 11,021.

2.22 Clinical Data

Clinical data is submitted to the UKMSR via clinical system or by eCRF:

- 8 Sites use the iMed MS Clinical System
- 2 Sites utilise another dedicated MS Clinical System
- 14 return an Electronic Case Return Form (eCRF)

At the time of analysis, iMed sites had submitted 2,306 patients, of these 163 had missing or invalid data, leaving 2,143 valid records. For the eCRF sites 888 records were submitted, when tested against the same criteria as for iMed 860 remained, leaving a combined total of 3,003. [Figure 2]

As of March 2017, there were 6,092 clinical informed consents, not all consented patients proceed online to provide data – only 4,053 of these consented individuals even provided an email address.

2.23 Linked

Deterministic matching required patients to exist in the portal and clinical datasets by study ID. For the 860 eCRF records 9 were missing variables leaving 851 valid records. For iMed - 2,705 records were checked using the same criteria leaving 1,776 (929 removed). For the portal, of the 11,021 entries checked for study ID with 9,619 removed, leaving 1402 records for analysis. Once linkage was made (allowing for nulls and incorrect values in diagnosis) there were 676 eligible records in both the clinical and portal data.

Figure 2: Consort diagram showing data selection criteria

3. Results

Data from the three sources: Portal, Clinical and Linked were compared and analysed.

3.1 Demographic Details

A total of 11,021 participants were included in this study 8,052 female and 2969 males (F:M 2.7:1). This population was subdivided between clinical and linked and more complete demography is described in Table 1.

	Clinical	Portal	Linked
	n=3,003	n=11,021	n=676
Age (mean)	48.8 ± 11.9	52.3 ± 11.7	48.3 ±11.3
Age at diagnosis	37.4 ± 10.6	39.3 ± 10.2	38.6 ± 10.6

Gender (female)	2,178 (75.2%)	8,052 (73.1%)	493 (72.9%)
PPMS	198 (6.5%)	1,514 (13.7%)	51(7.5%)
RRMS	2,564 (85.3)	7,408 (67.2%)	567(83.8%)
SPMS	122 (4.0%)	839 (7.6%)	21(3.1%)
Other	119(3.9%)	1,260 (11.4%)	37(5.4%)

Table 1: UKMSR datasets compared by age, age at diagnosis and MS Type at diagnosis

Mean age at entry to the study for the portal population was 52.3 years (SD 11.9), for the purely clinical population 48.8 (SD 48.8) and for the linked set 48.3 (SD 11.3). Mean age at diagnosis for the clinical population 37.4 (SD 37.4) and the portal 39.3 (SD 39.3), the overlapping linked population having a mean of 38.6 (SD 10.6). At time of capture the 85.3% of the clinical population were recorded as having RRMS, the portal reported 67.2% as RRMS and the linked as 83.8%. 10.5% of the clinical population were diagnosed with a progressive form of the disease with 21.3% of the portal indicating the same. 10.6% of the linked participants had progressive MS.

Table 2 shows a more complete breakdown of the reported MS Types by gender and data source.

	Total Linked	Female %	Total Portal	Female%	Total Clinical	Female %
PPMS	50	52%	1514	52.4%	198	48.4%
RRMS	567	74.8%	7408	77.7%	2,521	74.6%
SPMS	21	76.2%	839	66.3%	122	66.3%
Other	37	72.9%	1260	74.4%	162	73.4%

Table 2: Gender distribution across the UKMSR

For distributions of gender by disease type. Females show in higher proportions than males across all disease types except PPMS. In the linked population 52% of PPMS are female, in the Portal 52.4% with only the male clinical population deviating from this trend at 51.6%.

3.1.2 Location

The location of UKMSR participants against the UK as a whole was carried out using Register and ONS data. English data matched most closely with the portal and clinical populations 75.63% of the entire portal population being resident in an English Lower Super Output area. Northern Irish and Welsh participants being over represented at 12.36% and 12.01% respectively. [Table 3]. The UKMSR currently has no clinical sites within Scotland.

Country	UKMSR Portal % population	UKMSR Clinical population %	Entire UK population (17) %
England	77.6	75.63	84.2
Northern Ireland	4.3	12.36	2.8

Wales	7.7	12.01	4.7
Scotland	10.43	0	8.2

Table 3: UKMSR population distribution compared to the UK general population

3.2 Validity of the data

Comparing the ages of both the portal and clinical data using the two-sample K-S test revealed the populations were for the most part drawn from different distributions. The null hypothesis was rejected in the comparison of the datasets for both the current ages (D = 0.131, $p << 0.01$) and the ages at diagnosis (D = 0.078, $p << 0.01$). As the D statistic is small, the overall difference in the age distributions is minimal [Figures 3 and 4]. When stratifying by age and RRMS disease type alone, the null hypothesis is still rejected, but by a much smaller margin (D = 0.131, $p < 0.01$). Figure 3 shows the overall kernel density of the UKMSR clinical and portal populations for age at diagnosis, with Figure 4 highlighting the same data for people with relapsing remitting disease.

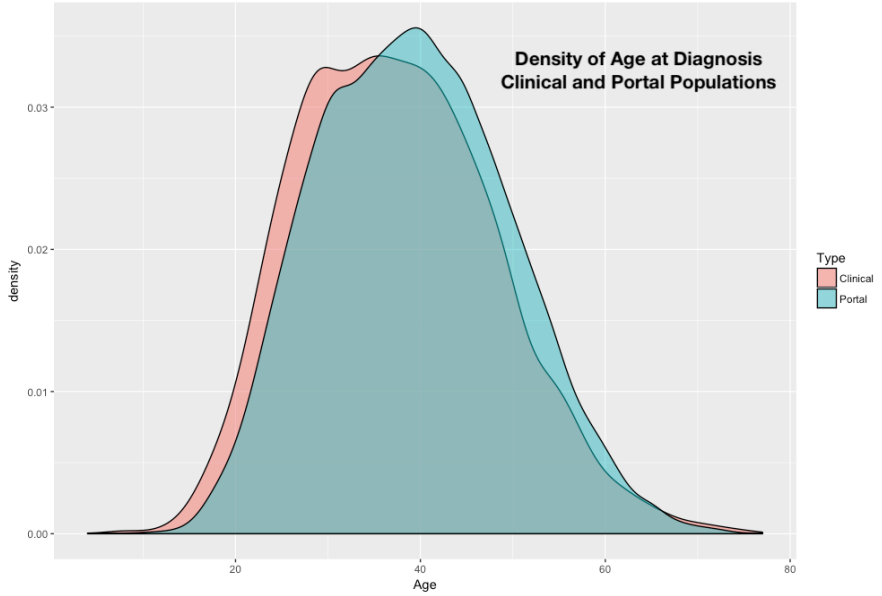

Figure 3 Kernel density of age at diagnosis, Portal and Clinical Population

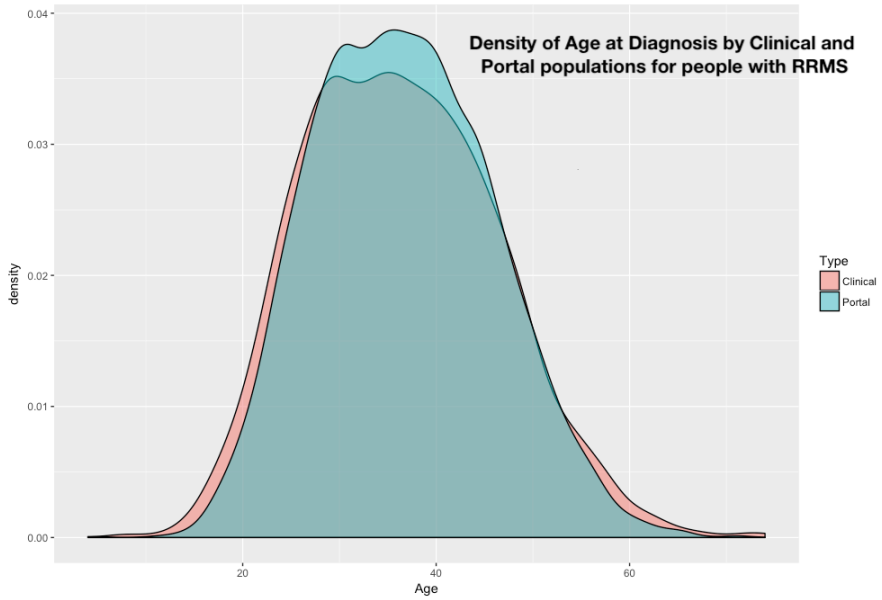

Figure 3 Kernel density of current age at diagnosis, Portal and Clinical Population for people with RRMS

Both the clinical and linked data show RRMS at diagnosis as >70% with the portal population at 67% being only slightly under this threshold. There are small differences between the populations declaring SPMS, 'Other' and PPMS as observed on the portal.

3.3 Quality of Self-Reported Data

Table 4 shows those data supplied by linked participants, compared to their clinical data. There are 99% accuracy in matches for gender and date of birth. The 1% likely to be transcription errors. Over 80% of clinicians and PwMS agree on the type of MS at diagnosis. 62% of participants remember the year of diagnosis exactly and this percentage rises to 81.7%

when a 1-year margin of error is allowed. For a population that have a mean of 13 years since diagnosis this is statistically insignificant.

Fields – Linked Clinical / Portal	Count	%
Total Linked Records	667	100
Matched Gender	664	99.55
Matching Date of Birth	652	97.7
Matched Year of Diagnosis	430	64.47
Date supplied within 1 year of Clinician Diagnosis	116	17.39
Date supplied within 2 years of Clinician Diagnosis	37	5.55
Date supplied within 3 years of Clinician Diagnosis	20	3.00
Date more than 3 years of clinical diagnosis date	37	5.55
Errors in Diagnosis Date	27	4.05
MS Type given exact match with current MS Type from clinician	549	82.31

Table 4: Quality of self-reported variables against clinically reported data.

4. Discussion

The objective of this study was to assess if the online element of the MS Register could be treated as a valid cohort of people with MS in the absence of supporting clinical validation. The data presented above, particularly the similarities in portal and clinical data ($D = 0.078$, $p << 0.01$) supports this assumption, and will allow researchers to work with Register data sets with more confidence than was previously possible. It has however become clear during this analysis that there are more patients with relapsing remitting disease in the clinical data set than has been seen in other studies (Clinical 85.3%, Portal 67.2%). Previous MS registry studies (18), (19) the LORSEP Registry (17) and the Atlas of MS (20) have also indicated an RRMS population of more than 80% in newly diagnosed patients. This is clearly in agreement with the large proportion of RRMS diagnosed patients coming to the UK MS Register from our clinical sites. In 2015 the Neuroinflammatory service in Cardiff defined their PPMS cohort as being 11% of their total population (21). In the MS Clinical data, progressive data combined has a total of 10.4% and supports the skew to RRMS further.

All other data are consistent with previous studies, there are more males with PPMS (22) but more females with other disease types. (23). The clinical group is on average 10 years younger than the portal - though well within the standard deviation, MS is a disease that is classically diagnosed in the mid to late 30's (24) and the date of diagnosis for both clinical and portal are within 0.07% of each other. Age at diagnosis of the portal population aligns well with other studies. (25), (19). One advantage to having a younger population overall is that they are more likely to make use of the internet to carry out PRO reporting (26). Internet usage is decreased in chronic disease (27) though overall UK Internet use amongst all age groups is increasing (17) with only 10.2% of the population never having accessed the Internet. This

may account for some of the disparity in the linked figures, but overall makes the collection of data using this methodology useful.

An interesting aspect of the analysis was the comparison of the linked cohort's responses against their clinical data with 81.7 % of patients remembering their diagnosis date to within 1 year of accuracy. This strongly indicates that participants are reliable narrators of their disease which is essential for the utility and validity of the other PROMs in the MS Register.

The capture of 'real world' / observational data rather than specific trial level data clearly brings with it a number of issues not typically seen in other cohort studies. The design of the UKMSR was to work with existing NHS clinical systems primarily to validate participants diagnoses whilst being mindful of the extra demands that data collections could have on busy NHS Staff (28).

Limitations

Other notable limitations include, the sample size of the linked population and the data quality of some captured variables from all sources – perhaps to be expected given the opportunistic capture methodology. The lack of clinical data from Scotland, is unfortunate but explained by the presence of the Scottish MS Register (16). These issues are all being addressed with more validation required by clinical sites submitting data. More stringent requirements for the self-entered data fields via the portal have been set.

The MS Register only captures age at confirmed diagnosis, rather than age at onset as this is a date that is more readily available within clinical records. The definition varied in the literature and in some other studies (29), (30) but it is a variable that could be added to the clinical minimum dataset. Age at onset is however captured from participants via the portal.

The recruitment of 'self-declared' people with MS via the internet is an inherent limitation, due to selection bias. This is slightly balanced by having 'general' recruitment at a variety of neurology centres across the UK – with the caveats about increased numbers of RRMS noted. To provide more balance to this, the UKMSR is now attempting to recruit more community care NHS Trusts and to encourage exiting sites to make more use of the postal consent methodology to capture those patients that may no longer attend clinics.

5. Conclusions and future work

This validation work demonstrates that in comparison to a clinical population there is representative sample emerging in the online portal population. This paper is a firm first step in being able to treat the online cohort of people with MS as a valid one and underpins the outcomes-based research that has already been carried out via the UKMSR.

For the future the UKMSR will improve the quality and quantity of data collected from clinical systems and the internet by tightening the validation requirements of data entry. More linked data sets will be sought in order to increase data linkage and overall data quality. A potential

example of this would be by comparison with another well characterised prevalent population, such as the Scottish incidence register.

Clinically, capture of data will be improved from the NHS by transitioning to a new eCRF system that is easier to access from within the NHS and provides instant feedback on collected clinical data. Similar improvements on the portal collection methodology are planned and new methods of data capture will be provisioned, such as smartphone apps to collect participant activity data as a potentially concordant outcome measure with EDSS (31) .

Around 30% of patients on average recruited via clinical sites that then go on to use UKMSR online. We therefore need to recruit more patients clinically, but also from sources more varied than secondary care specialist treatment centres, perhaps extending to more community trusts and rehabilitation teams. Additionally, we are examining the application of Natural Language Processing techniques as a data capture methodology.

The methodology of data capture from the internet, clinical systems and routine data has proven to be useful and crucially it is patient centred. Having established the validity of the cohort it becomes possible to make use of this research in a variety of ways. The first will be selecting subsets of the cohort that may be appropriate for clinical trials and making sure the broader dataset is fit for purpose. Of additional interest will be testing the whole cohort with novel online outcome measures that would be difficult to test on less well-characterised or smaller clinical cohorts. For example, how a web-based participant supplied EDSS score compares to a formal clinically supplied one.

Acknowledgements

The UKMSR is in receipt of a grant from the MS Society. Award 64

6. References

1. Brownlee WJ, Hardy TA, Fazekas F, Miller DH. Diagnosis of multiple sclerosis: progress and challenges. The Lancet. 2017 Apr;389(10076):1336–46.

2. Leray E, Yaouanq J, Le Page E, Coustans M, Laplaud D, Oger J, et al. Evidence for a two-stage disability progression in multiple sclerosis. Brain. 2010 Jul 1;133(7):1900–13.

3. Thompson AJ. Challenge of progressive multiple sclerosis therapy: Curr Opin Neurol. 2017 Jun;30(3):237–40.

4. Ford DV, Jones KH, Verplancke J-P, Lyons RA, John G, Brown G, et al. The SAIL Databank: building a national architecture for e-health research and evaluation. BMC Health Serv Res. 2009;9(1):157.

5. Flachenecker P, Stuke K. National MS registries. J Neurol. 2008 Dec;255(S6):102–8.

6. Sayers A, Ben-Shlomo Y, Blom AW, Steele F. Probabilistic record linkage. Int J Epidemiol. 2016 Jun;45(3):954–64.

7. Health and Social Care Information Centre. NHS ICD-10 5th Edition XML data files releases [Internet]. NHS Classifications ICD10 5th Edition. 2015 [cited 2016 Nov 10]. Available from: https://isd.hscic.gov.uk/trud3/user/guest/group/0/pack/28/subpack/259/releases

8. NHS Digital. UK Read Code [Internet]. UK Read Code. [cited 2016 Nov 10]. Available from: https://data.gov.uk/dataset/uk-read-code/resource/3314d8a4-ef2c-4a00-a7d7-ddec14b88519

9. Marrie RA, Cutter G, Tyry T, Campagnolo D, Vollmer T. Validation of the NARCOMS registry: diagnosis. Mult Scler. 2007 Feb 9;13(6):770–5.

10. UK MS Register. UK MS Regsiter Terms and Conditions [Internet]. UK MS Register, terms and conditions. 2012 [cited 2016 Nov 16]. Available from: https://www.ukmsregister.org/Portal/TAndC

11. Zigmond, A, Snaith, R.P. The Hospital Anxiety and Depression Scale. Acta Psychiatr Scaninavica. 1983;67(6):361–70.

12. Brooks R. EuroQol: the current state of play. Health Policy. 1996 Jul;37(1):53–72.

13. Hobart J. The Multiple Sclerosis Impact Scale (MSIS-29): A new patient-based outcome measure. Brain. 2001 May 1;124(5):962–73.

14. Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria. Ann Neurol. 2011 Feb;69(2):292–302.

15. R Development Team. R: a language and environment for statistical computing [Internet]. 2011. Available from: https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing

16. Ford DV, Jones KH, Middleton RM, Lockhart-Jones H, Maramba ID, Noble GJ, et al. The feasibility of collecting information from people with Multiple Sclerosis for the UK MS Register via a web portal: characterising a cohort of people with MS. BMC Med Inform Decis Mak. 2012;12(1):73.

17. ONS. Population Estimates for UK, England and Wales, Scotland and Northern Ireland: mid-2015 [Internet]. Statistical Bulletin. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2015#uk-population-reached-651-million-in-mid-2015

18. Boiko A, Vorobeychik G, Paty D, Devonshire V, Sadovnick D, the UBC MS Clinic Neurologists. Early onset multiple sclerosis: A longitudinal study. Neurology. 2002 Oct 8;59(7):1006–10.

19. Sumelahti M-L, Holmberg MHA, Murtonen A, Huhtala H, Elovaara I. Increasing Incidence in Relapsing-Remitting MS and High Rates among Young Women in Finland: A Thirty-Year Follow-Up. Mult Scler Int. 2014;2014:1–8.

20. MSIF. Atlas of MS [Internet]. Atlas of MS. 2013 [cited 2016 Nov 7]. Available from: https://www.msif.org/about-us/advocacy/atlas/

21. Harding KE, Wardle M, Moore P, Tomassini V, Pickersgill T, Ben-Shlomo Y, et al. Modelling the natural history of primary progressive multiple sclerosis. J Neurol Neurosurg Psychiatry. 2015 Jan;86(1):13–9.

22. Miller DH, Leary SM. Primary-progressive multiple sclerosis. Lancet Neurol. 2007 Oct;6(10):903–12.

23. Ribbons KA, McElduff P, Boz C, Trojano M, Izquierdo G, Duquette P, et al. Male Sex Is Independently Associated with Faster Disability Accumulation in Relapse-Onset MS but Not in Primary Progressive MS. Aktas O, editor. PLOS ONE. 2015 Jun 5;10(6):e0122686.

24. Paulus A, Hussack S, Kugler J. Multiple Sklerose – Ergebnisse einer Befragung der Mitglieder des DMSG-Landesverbandes e.V. in Sachsen-Anhalt: Durch welche Faktoren wird die Diagnosedauer beeinflusst? Fortschritte Neurol · Psychiatr. 2016 Aug 29;84(08):487–93.

25. Mackenzie IS, Morant SV, Bloomfield GA, MacDonald TM, O'Riordan J. Incidence and prevalence of multiple sclerosis in the UK 1990-2010: a descriptive study in the General Practice Research Database. J Neurol Neurosurg Psychiatry. 2014 Jan 1;85(1):76–84.

26. Cecil Prescott. Internet users in the UK: 2016 [Internet]. Statistical Bulletin. 2016 [cited 2017 Apr 26]. Available from: https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/internetusers/2016

27. Fox S, Purcell, Kristen. Chronic Disease and the Internet [Internet]. Pew Research Center. 2010 [cited 2016 Nov 9]. Available from: http://www.pewinternet.org/2010/03/24/chronic-disease-and-the-internet/

28. Waterson P. Health information technology and sociotechnical systems: A progress report on recent developments within the UK National Health Service (NHS). Appl Ergon. 2014 Mar;45(2):150–61.

29. Kremenchutzky M. The natural history of multiple sclerosis: a geographically based study 9: Observations on the progressive phase of the disease. Brain. 2006 Jan 6;129(3):584–94.

30. Leray E, Moreau T, Fromont A, Edan G. Epidemiology of multiple sclerosis. Rev Neurol (Paris). 2016 Jan;172(1):3–13.

31. Motl RW, Pilutti LA, Learmonth YC, Goldman MD, Brown T. Clinical Importance of Steps Taken per Day among Persons with Multiple Sclerosis. Villoslada P, editor. PLoS ONE. 2013 Sep 4;8(9):e73247.

Appendix 1

List of clinical sites that contributed data to this study

ABMU NHS Trust Morriston Hospital
Nottingham Hospitals NHS Trust, Queens Medical Centre
UCL London Hospitals NHS Trust National Hospital for Neurology and Neurosurgery
Belfast Health and Social Care Trust, Belfast City Hospital
Barts Health NHS Trust, Royal London Hospital
Shrewsbury and Telford NHS Trust, Royal Shrewsbury Hospital
Basildon and Thurrock University Hospitals NHS Foundation Trust, Basildon Hospital
University Hospital Southampton NHS Foundation Trust. Southampton General Hospital
Oxford University Hospitals NHS Foundation Trust John Radcliffe Hospital
Poole Hospital NHS Foundation Trust, Poole Hospital
Brighton and Sussex University Hospitals NHS Trust , Princess Royal Hospital
Royal Free London NHS Foundation Trust, Royal Free Hospital
Hampshire Hospitals NHS Foundation Trust, Basingstoke and North Hampshire Hospital
Salford Royal NHS Foundation Trust, Salford Royal Hospital
Northampton NHS Trust, Northampton General Hospital
Royal Devon and Exeter NHS Foundation Trust, Royal Devon and Exeter Hospital
Luton and Dunstable Hospital NHS Foundation Trust, Luton Hospital
Frimley Park NHS Foundation Trust, Frimley Park Hospital
Barking, Havering And Redbridge Hospitals NHS Trust, Queens Hospital
Mid Yorkshire Hospitals NHS Trust, Pinderfields General Hospital
Royal Cornwall Hospitals NHS Trust, Royal Cornwall Hospitals NHS Trust
Southend University Hospital NHS Foundation Trust, Southend Hospital

Appendix 2 UK MS Register Minimum Dataset
IP CAG

MS REGISTER

Minimum Data Set
Version 6.3

Name & Address

Date	/ /

Study ID							

TIER ONE (all fields to be completed)

MS Type now				
RR	SP	PP	Other	
Conversion to SP (if applicable)				

Date of Diagnosis	/ /	/	MS Type at Diagnosis	RR	SP	PP	Other
Date of onset	/ /						

No. of Relapses (RR only) (since last visit/year)		**Severity:** (circle)	Mild	Moderate	Severe

PAST Disease modifying Treatment (circle)					**Date Started**	/ /
Alemtuzumab	Avonex	Betaferon	Copaxone	Extavia	Fingolimod	Mitoxantrone
Natalizumab	Ocrelizumab	Rebif	Tecfidera	Teriflunomide	None	
Date Stopped	/ /	**Reason** (circle)	Lack of efficacy / Side Effects / Other			

PRESENT Disease modifying Treatment (circle)					**Date Started**	/ /
Alemtuzumab	Avonex	Betaferon	Copaxone	Extavia	Fingolimod	Mitoxantrone
Natalizumab	Ocrelizumab	Rebif	Tecfidera	Teriflunomide	None	

Current EDSS Score (1-10)		**Date EDSS Taken**	/ /	
Walking range : time and distance in meters	Time = _____ M = _____	Self-estimated	Trundle wheel	Treadmill

TIER TWO (to be completed IF patient is unlikely to register online)

Patient is: (tick)

Pregnant	Y	N

Smoker	Y	N
No. Per Day		
Smoked since:		

Onset Localisation (circle)	
Spinal	Cortex
Visual	Cerebellar/brainstem

Onset Symptoms (circle)		
Vision	Motor	Sensory
Coordination	Bowel/Bladder	Fatigue
Cognitive	Encephalopathy	Other

Person completing form: _____