This is an author produced version of a paper published in:
Multiple Sclerosis and Related Disorders

Cronfa URL for this paper:
http://cronfa.swan.ac.uk/Record/cronfa40629

Paper:
http://dx.doi.org/10.1016/j.msard.2018.05.015

This item is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence. Copies of full text items may be used or reproduced in any format or medium, without prior permission for personal research or study, educational or non-commercial purposes only. The copyright for any work remains with the original author unless otherwise specified. The full-text must not be sold in any format or medium without the formal permission of the copyright holder.

Permission for multiple reproductions should be obtained from the original author.

Authors are personally responsible for adhering to copyright and publisher restrictions when uploading content to the repository.

http://www.swansea.ac.uk/library/researchsupport/ris-support/
Validating the Portal Population of the United Kingdom Multiple Sclerosis Register


PII: S2211-0348(18)30176-7
DOI: 10.1016/j.msard.2018.05.015
Reference: MSARD 853

To appear in: Multiple Sclerosis and Related Disorders

Received date: 2 February 2018
Revised date: 17 May 2018
Accepted date: 22 May 2018


This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Highlights

- Validation of a self declared MS Population with a clinically diagnosed one.
- We analysed the UK MS Register for a number of key characteristics. The internet (n=11,021) and clinical (n=3,003) populations were studied for key shared epidemiology. 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)
- There is a representative population of people with MS in the online element of the UK MS – The MS Register can be treated as a valid instrument for Clinical and PRO research
- Kernel Density of the populations for Age can be seen below:
Title: Validating the Portal Population of the United Kingdom Multiple Sclerosis Register

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

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

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

![Venn diagram showing the linked population as an intersection of the internet participants with consented patients from a clinical site.](image)

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.
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.

<table>
<thead>
<tr>
<th></th>
<th>Clinical</th>
<th>Portal</th>
<th>Linked</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>n=3,003</td>
<td>n=11,021</td>
<td>n=676</td>
</tr>
<tr>
<td>Age (mean)</td>
<td>48.8 ± 11.9</td>
<td>52.3 ± 11.7</td>
<td>48.3 ±11.3</td>
</tr>
<tr>
<td>Age at diagnosis</td>
<td>37.4 ± 10.6</td>
<td>39.3 ± 10.2</td>
<td>38.6 ± 10.6</td>
</tr>
</tbody>
</table>
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 1 shows a more complete breakdown of the reported MS Types by gender and data source.

<table>
<thead>
<tr>
<th>Gender (female)</th>
<th>Total Linked</th>
<th>Total Portal</th>
<th>Total Clinical</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPMS</td>
<td>2,178 (75.2%)</td>
<td>8,052 (73.1%)</td>
<td>493 (72.9%)</td>
<td>198 (6.5%)</td>
</tr>
<tr>
<td>RRMS</td>
<td>2,564 (85.3%)</td>
<td>7,408 (67.2%)</td>
<td>567 (83.8%)</td>
<td>1,514 (13.7%)</td>
</tr>
<tr>
<td>SPMS</td>
<td>122 (4.0%)</td>
<td>839 (7.6%)</td>
<td>21 (3.1%)</td>
<td>51 (7.5%)</td>
</tr>
<tr>
<td>Other</td>
<td>119 (3.9%)</td>
<td>1,260 (11.4%)</td>
<td>37 (5.4%)</td>
<td>37 (5.4%)</td>
</tr>
</tbody>
</table>

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

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.

<table>
<thead>
<tr>
<th>Country</th>
<th>UKMSR Portal % population</th>
<th>UKMSR Clinical population %</th>
<th>Entire UK population (17) %</th>
</tr>
</thead>
<tbody>
<tr>
<td>England</td>
<td>77.6</td>
<td>75.63</td>
<td>84.2</td>
</tr>
<tr>
<td>Northern Ireland</td>
<td>4.3</td>
<td>12.36</td>
<td>2.8</td>
</tr>
</tbody>
</table>
Table 3: UKMSR population distribution compared to the UK general population

<table>
<thead>
<tr>
<th></th>
<th>Wales</th>
<th>12.01</th>
<th>4.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scotland</td>
<td>10.43</td>
<td>0</td>
<td>8.2</td>
</tr>
</tbody>
</table>

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.
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.

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

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
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

**Conflict of Interest / Funding Source for Validating the Portal Population of the UK MS Register**

The authors declare that there is no conflict of interest regarding the publication of this article. There are no relationships to people or organisations that have biased this work.

This work has not been published previously and has not been submitted concurrently to any other journals

Funding for this research has been provided by the MS Society. Award 64h
6. References


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


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


Appendix 1

List of clinical sites that contributed data to this study

<table>
<thead>
<tr>
<th>Clinical Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABMU NHS Trust Morriston Hospital</td>
</tr>
<tr>
<td>Nottingham Hospitals NHS Trust, Queens Medical Centre</td>
</tr>
<tr>
<td>UCL London Hospitals NHS Trust National Hospital for Neurology and Neurosurgery</td>
</tr>
<tr>
<td>Belfast Health and Social Care Trust, Belfast City Hospital</td>
</tr>
<tr>
<td>Barts Health NHS Trust, Royal London Hospital</td>
</tr>
<tr>
<td>Shrewsbury and Telford NHS Trust, Royal Shrewsbury Hospital</td>
</tr>
<tr>
<td>Basildon and Thurrock University Hospitals NHS Foundation Trust, Basildon Hospital</td>
</tr>
<tr>
<td>University Hospital Southampton NHS Foundation Trust, Southampton General Hospital</td>
</tr>
<tr>
<td>Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital</td>
</tr>
<tr>
<td>Poole Hospital NHS Foundation Trust, Poole Hospital</td>
</tr>
<tr>
<td>Brighton and Sussex University Hospitals NHS Trust, Princess Royal Hospital</td>
</tr>
<tr>
<td>Royal Free London NHS Foundation Trust, Royal Free Hospital</td>
</tr>
<tr>
<td>Hampshire Hospitals NHS Foundation Trust, Basingstoke and North Hampshire Hospital</td>
</tr>
<tr>
<td>Salford Royal NHS Foundation Trust, Salford Royal Hospital</td>
</tr>
<tr>
<td>Northampton NHS Trust, Northampton General Hospital</td>
</tr>
<tr>
<td>Royal Devon and Exeter NHS Foundation Trust, Royal Devon and Exeter Hospital</td>
</tr>
<tr>
<td>Luton and Dunstable Hospital NHS Foundation Trust, Luton Hospital</td>
</tr>
<tr>
<td>Frimley Park NHS Foundation Trust, Frimley Park Hospital</td>
</tr>
<tr>
<td>Barking, Havering And Redbridge Hospitals NHS Trust, Queens Hospital</td>
</tr>
<tr>
<td>Mid Yorkshire Hospitals NHS Trust, Pinderfields General Hospital</td>
</tr>
<tr>
<td>Royal Cornwall Hospitals NHS Trust, Royal Cornwall Hospitals NHS Trust</td>
</tr>
<tr>
<td>Southend University Hospital NHS Foundation Trust, Southend Hospital</td>
</tr>
</tbody>
</table>
## Appendix 2 UK MS Register Minimum Dataset

### TIER ONE (all fields to be completed)

<table>
<thead>
<tr>
<th>MS Type now</th>
<th>Date of Diagnosis</th>
<th>MS Type at Diagnosis</th>
<th>Date of onset</th>
</tr>
</thead>
<tbody>
<tr>
<td>RR</td>
<td>/ /</td>
<td>/ /</td>
<td></td>
</tr>
<tr>
<td>SP</td>
<td></td>
<td>RR</td>
<td></td>
</tr>
<tr>
<td>PP</td>
<td></td>
<td>SP</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td>PP</td>
<td></td>
</tr>
</tbody>
</table>

### Conversion to SP (if applicable)

<table>
<thead>
<tr>
<th>No. of Relapses (RR only)</th>
<th>Severity (circle)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(since last visit/year)</td>
<td>Mild</td>
</tr>
</tbody>
</table>

### 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

<table>
<thead>
<tr>
<th>Time =</th>
<th>Self-estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>M =</td>
<td>Trundle wheel</td>
</tr>
<tr>
<td></td>
<td>Treadmill</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Patient is: (tick)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pregnant</td>
</tr>
<tr>
<td>Smoker</td>
</tr>
</tbody>
</table>

### Onset Localisation (circle)

- Spinal
- Cortex
- Visual
- Cerebellar/brainstem

### Onset Symptoms (circle)

- Vision
- Motor
- Sensory
- Coordination
- Bowel/Bladder
- Fatigue
- Cognitive
- Encephalopathy
- Other

### Person completing form: ________________________________

UK MS Register MDS 23/08/17