Lessons Learned in Identifying Relapsing-Remitting Multiple Sclerosis in US Integrated Delivery Network Health Care Claims and Electronic Health Record Data

Hoa Van Le, MD, PhD (Presenter) PAREXEL International ISPOR 22nd Annual International Meeting, Boston, MA

May 23, 2017
Authors

Hoa V. Le, MD PhD\textsuperscript{1}; Chi T. L. Truong, MD PhD\textsuperscript{2}; Aaron W. C. Kamauu, MD MS MPH\textsuperscript{3}; John R. Holmén, PhD MS\textsuperscript{4}; Christopher L. Fillmore, MD MS\textsuperscript{4}; Meritxell Sabidó-Espin, MD, MPH, PhD\textsuperscript{5}; Schiffon L. Wong, MPH\textsuperscript{6}

\textsuperscript{1}PAREXEL International, Durham, NC, USA
\textsuperscript{2}MedCodeWorld, Mississauga, ON, Canada
\textsuperscript{3}Anolinx LLC, Salt Lake City, UT, USA
\textsuperscript{4}Intermountain Healthcare, Murray, UT, USA
\textsuperscript{5}Merck KGaA, Darmstadt, Germany
\textsuperscript{6}EMD Serono, Inc., Billerica, MA, USA
Disclosures

• This study was funded by EMD Serono / Merck KGaA
• HVL is an employee of PAREXEL and a stockholder of GlaxoSmithKline; and was a merit scholarship recipient (Harry Guess – Merck Award) from Merck
• CTLT is an employee of MedCodeWorld
• AWCK received a research grant from Merck KGaA
• JRH and CLF have nothing to disclose
• MSE is an employee of Merck KGaA
• SLW is an employee of EMD Serono
Background

- Real world data can be beneficial for conducting observational studies in relapsing-remitting multiple sclerosis (RRMS) that focus on comparative effectiveness and safety of medical products.

- **Algorithms** may be used to define **patient cases, outcomes, or events** based on diagnoses, procedures and treatments.

- **RRMS algorithm development and validation are essential** to ensure high quality research and accurate results using real world data.
Multiple Sclerosis

Healthy nerve

Nerve demyelination in multiple sclerosis
Multiple Sclerosis Subtypes

- In the US healthcare databases, four MS subtypes are recorded under one ICD-9 code 340

**Primary-Progressive (PPMS)**

**Clinically Isolated Syndrome (CIS)**
1st episode of inflammatory demyelination

**Relapsing-Remitting (RRMS)**

**Secondary-Progressive (SPMS)**

Lublin et al. 2014
Study Objective

- To develop and validate claims- and Electronic Health Record (EHR)-based algorithms to identify RRMS in a US Integrated Delivery Network (IDN) healthcare system
Methods

• **Study period:** 1 January 2010 – 31 December 2014

• **Study population**
  – Age ≥ 18
  – At least 1 Multiple Sclerosis (MS) Diagnosis OR MS Disease-Modifying Therapy (DMT) and history of MS diagnosis
  – No other demyelinating diseases, no pregnancy
  – 12-month healthcare coverage

• **Index date:** date of first MS diagnosis or DMT in the study period
RRMS Study Population: Claims-based

**INCLUSIONS**
Multiple sclerosis (MS) patient identification by combinations of:
- **MS diagnosis**
- Specific **MS symptoms** during a neurology visit
- Use of disease-modifying therapy (DMT), or
  - Brain/spinal magnetic resonance imaging (MRI)

**EXCLUSIONS**
Patients with **progressive disease** were excluded by one of the following options:

**Option A:** Medications often used for progressive disease

**Option B:** Change of Expanded Disability Status Scale (EDSS) scores based on a conversion of Kurtzke Functional Systems Scores (KFSS) into ICD-9-CM

**Option C:** Pattern of supportive therapy use (e.g., nursing home, home health, selected rehabilitation/durable medical equipment [DME]) over 12 months*

**RRMS COHORTS**
- Cohort A
- Cohort B
- Cohort C

*Adapted from Gilden et al. 2011
Patients with ≥ 1 clinical document with mention of MS AND ≥ 1 NLP-based mention of any of the terms/phrases for clinician-documented diagnosis of RRMS in a clinical note during the study period EXCLUDING ≥ 1 NLP-based mention of any of the terms/phrases for clinician-documented diagnosis of progressive MS in a clinical note during the study period.

RRMS Cohort*

* Clinically stable based on progression/ not-progressive
Algorithm Evaluation

- Natural Language Processing (NLP)-based manual medical chart reviews
- Random sample medical chart reviews were the “gold standard” for algorithm validation
- Positive predictive value (PPV) calculations
- Sensitivity analyses
MS identification by Claims-based Algorithm: Inclusion Contributions

Total number of patients who met inclusion criteria (n=2,960)

- ≥1 Diagnosis + 1 of 5 following
  - ≥1 MS-indicated DMT (n=1545)
  - ≥1 MS specific symptom therapy during a neurology visit (n=1,483)
  - ≥1 MS specific symptom therapy during a neurology visit ≥ 30 days apart (n=1,112)
  - ≥1 Brain or Spinal MRI before index (n=872)
  - ≥ 1 ICD-9 code 378.86 at least 30 days apart (n=5)

- ≥1 DMT + diagnosis history + 1 of 3 following
  - ≥1 Brain or Spinal MRI (n=714)
  - ≥ 1 MS specific symptom therapy during a neurology visit (n=479)
  - ≥1 MS specific symptom during a neurology visit (n=275)
Claims-based Algorithm Exclusions: Progressive MS Identification

**Total 689 Progressive MS patients excluded based on one of the 3 options**

**Option A (9% = 60/689)**
- Medications often used for progressive disease (mitoxantrone, cyclophosphamide, or methotrexate)

**Option B (88% = 608/689)**
- Disease progression based on a specific change of EDSS scores in the last 12 months of the patient's most recent year of care coverage after index date and during the study period

**Option C (6% = 44/689)**
- At least 12 months of recorded MS history and one of the following:
  - At least 10 of the last 12 months at the exacerbation level
  - The last 12 months at the plateau/stable level with a final therapy type of nursing home, home health, selected rehabilitation/DME.

*based on EDSS scores

**Total number of Clinically Stable RRMS* patients who met inclusion criteria and excluding patients with progressive disease (n=2,271)**
Top 3 PPV for Claims-based Algorithm

- **Certain (n=56)**
  - PPV$_1$ = 88.6%
  - 95% CI: 75.4-95.1%
  - Positive
  - Negative

- **Likely (n=2)**
  - PPV$_2$ = 87.9%
  - 95% CI: 77.0-94.3%
  - Positive
  - Negative

- **Possible (n=4)**
  - PPV$_3$ = 87.5%
  - 95% CI: 78.3-94.1%
  - Positive
  - Negative

- **No (n=8)**
  - Certain (n=56)
  - Likely (n=2)
  - Possible (n=4)
  - No (n=8)
  - Unknown (n=34)

- **Unknown (n=34)**
  - Certain (n=56)
  - Likely (n=2)
  - Possible (n=4)
  - No (n=8)
  - Unknown (n=34)

-ISPOR May 23, 2017-
**EHR-based Algorithm: Number of patients meeting criteria**

Patients with $\geq 1$ clinical document with mention of **MS**

__AND__

$\geq 1$ Natural Language Processing (NLP)-based mention of any of the terms/phrases for clinician-documented **diagnosis of RRMS** in a clinical note during the study period

__EXCLUDING__

$\geq 1$ NLP-based mention of any of the terms/phrases for clinician-documented diagnosis of **progressive MS** in a clinical note during the study period

- **RRMS Cohort*:**
  - $N=837$
  - $N=4,623$
  - $N=153$
  - $N=990$

*Clinically stable based on progression/ not-progressive*
EHR-based Algorithm

Word Cloud - Top 100 terms
EHR-based Algorithm
Search terms and % hit of 62,909 documents

multiple'
sclerosis'
relapsing'
remitting'
progressive'
subtype'
RRMS'
multiple sclerosis'
relapsing remitting'
NEAR( (multiple, sclerosis) , 3, TRUE ) '
NEAR( (relapsing, remitting) , 4, FALSE ) '
NEAR( (relapsing, remitting, multiple, sclerosis) , 12, FALSE ) '
NEAR( (multiple, sclerosis, relapsing, remitting, subtype ) , 12, FALSE ) '
NEAR( (remittent, progressive , multiple, sclerosis ) , 12, FALSE ) '
NEAR( (multiple, sclerosis, relapsing, remitting, type) , 12, FALSE ) '
NEAR( ((not), multiple, sclerosis) , 20, FALSE ) '
NEAR( (unlikely, multiple, sclerosis) , 15, FALSE ) '

97%
96%
4.7%
4.2%
5.6%
0.055%
0.098%
94.8%
2.79%
95.2%
4%
2.8%
0.023%
0.003%
0.063%
0.95%
0.12%
# EHR-based Algorithm Search Terms

## RRMS Terms

<table>
<thead>
<tr>
<th>Term</th>
<th># of unique patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>relapsing remitting</td>
<td>839</td>
</tr>
<tr>
<td>relapsing</td>
<td>970</td>
</tr>
<tr>
<td>remitting</td>
<td>862</td>
</tr>
</tbody>
</table>

## Progressive MS terms

<table>
<thead>
<tr>
<th>Term</th>
<th># of unique patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>contains(document_text,' NEAR( (progressive, multiple, sclerosis), 6, FALSE ) ',18)&gt; 0</td>
<td>153</td>
</tr>
<tr>
<td>contains(document_text,'progressive', 5)&gt;0 Not used: Using this term proved too broad, resulting in most as false positives</td>
<td>522</td>
</tr>
</tbody>
</table>
PPV Options for EHR-based Algorithm

Certain (n=107)  Likely (n=0)  Possible (n=0)  No (n=1)  Unknown (n=3)

**Positive**
- PPV=99.1%
- 95% CI: 94.2-100%

**Negative**
- PPV=99.1%
- 95% CI: 94.4-100%

**Unknown**
- PPV=96.4%
- 95% CI: 90.5-98.8%

*splitting Unknowns: 85% RRMS vs 15% non-RRMS*
• **High Positive Predictive Value** (PPV) may be expected due to the high prevalence of RRMS among MS patients

• We chose an approach utilizing multiple criteria and several options to assess the contributions of each factor towards positively distinguish RRMS patients from
  – secondary progressive multiple sclerosis (SPMS) and
  – primary progressive multiple sclerosis (PPMS) patients
Discussion

• **Lack of documentation of MS subtype** in clinician’s documentation is challenging
  - *Unknown cases* during validation occurred when MS subtype was not explicitly included in the clinician's documentation
  - *One third of patients* selected for claims-based algorithm validation **did not** have subtype documented
  - *Only 21% of patients* identified by the EHR-based algorithm with a mention of MS **had a term for RRMS** without an excluding negation term

• NLP uses both **structure and unstructured** texts may improve the current EHR-based algorithm for unstructured clinical texts only

• **Traditional medical chart review** may enhance the NLP-based medical chart review
Conclusions

• Both the claims-based and EHR-based algorithms had excellent PPV for identifying RRMS among patients with documented MS subtypes

• Traditional medical chart reviews support the NLP-based chart reviews, particularly for patients without clinical notes of MS subtypes

• The claims-based and EHR-based algorithms to identify RRMS and NLP-based chart reviews are promising methods for future research.
Acknowledgements

Colleagues and Contributors from:

- PAREXEL International
- EMD Serono/ Merck KGaA
- Intermountain Healthcare
- Anolinx
- MedCodeWorld
- Others
Questions?

Thank you!

Contacts:

• **Hoa Le**: hoa.le@parexel.com
• **Schiffon Wong**: Schiffon.wong@emdserono.com