Creating a comorbidity index with machine learning: Work in progress

Hanna Frank <hanna.frank@ubc.ca>
PhD Candidate, SPPH
July 10, 2024

Collaborators:

- Dr. Helen Tremlett
- Dr. Ruth Ann Marrie
- Dr. Mohsen Sadatsafavi
- Dr. Lisa Lix
- Dr. Kyla McKay
- Feng Zhu
- Fardowsa Yusuf
- Dr. Mohammad Ehsanul Karim

Funding:

- UBC four-year fellowship
- MS Canada Discovery Research Grant

Research Question

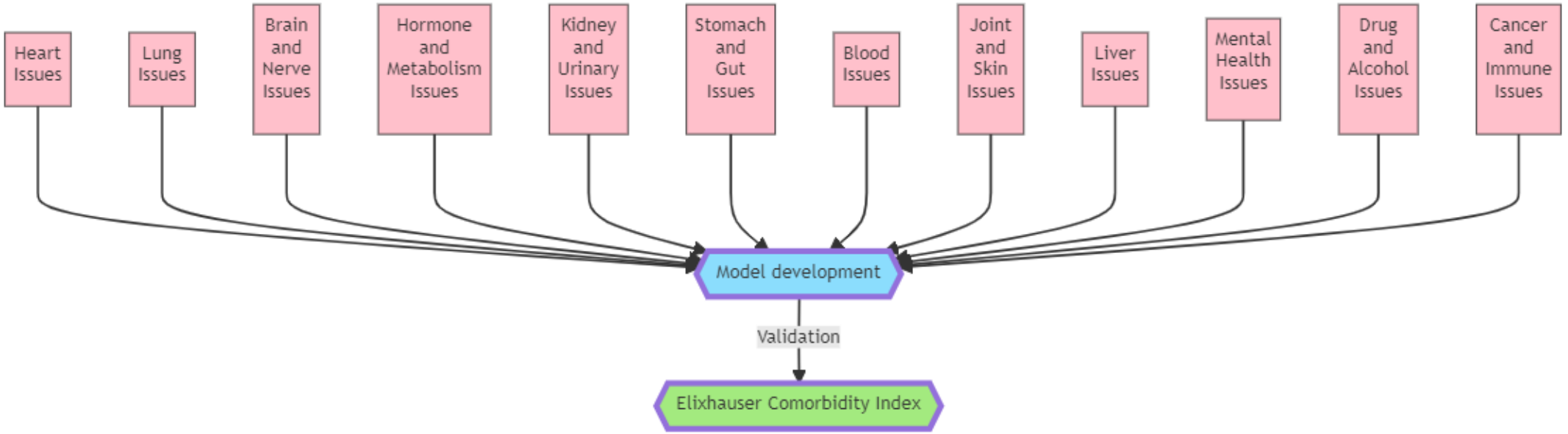
Overarching Epidemiological Question:

How can a **comorbidity index** be developed for people with multiple sclerosis (MS) using **health administrative data**?

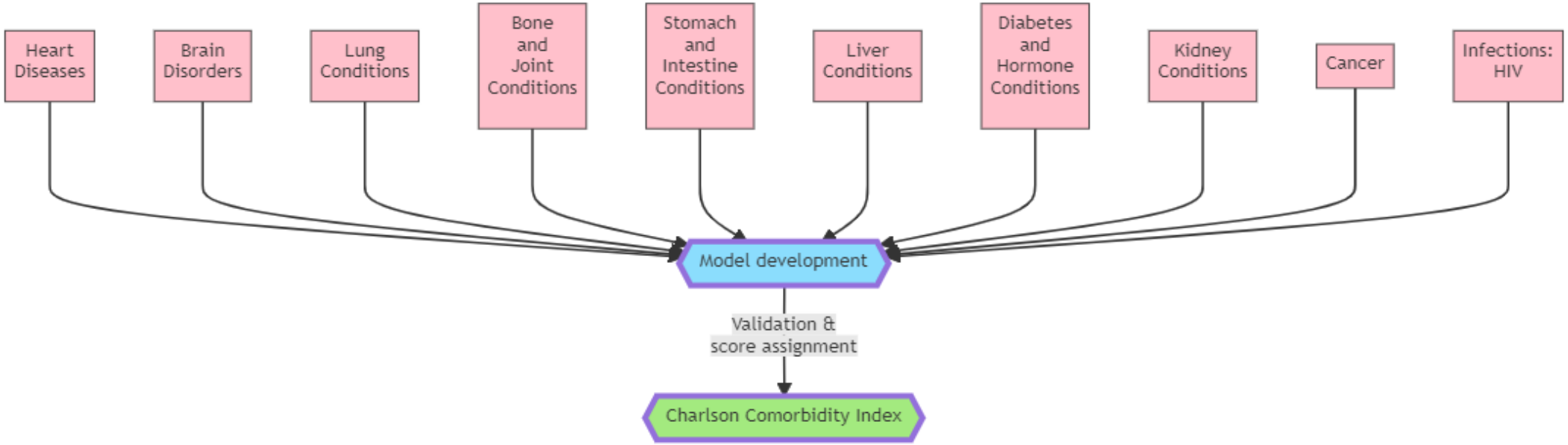
Methods Question:

How can **machine learning methods** be applied in the development of such a comorbidity index?

Currently available measures: Elixhauser Comorbidity Index (ECI)



Currently available measures: Charlson Comorbidity Index (CCI)



Currently available measures: Charlson Comorbidity Index (CCI)

Weights indicate the impact of each comorbidity on the outcome of interest.

Table 3. Weighted index of comorbidity

Assigned weights for diseases	Conditions
1	Myocardial infarct
	Congestive heart failure
	Peripheral vascular disease
	Cerebrovascular disease
	Dementia
	Chronic pulmonary disease
	Connective tissue disease
	Ulcer disease
	Mild liver disease
	Diabetes
2	Hemiplegia
	Moderate or severe renal disease
	Diabetes with end organ damage
	Any tumor
	Leukemia
	Lymphoma
3	Moderate or severe liver disease
6	Metastatic solid tumor
	AIDS

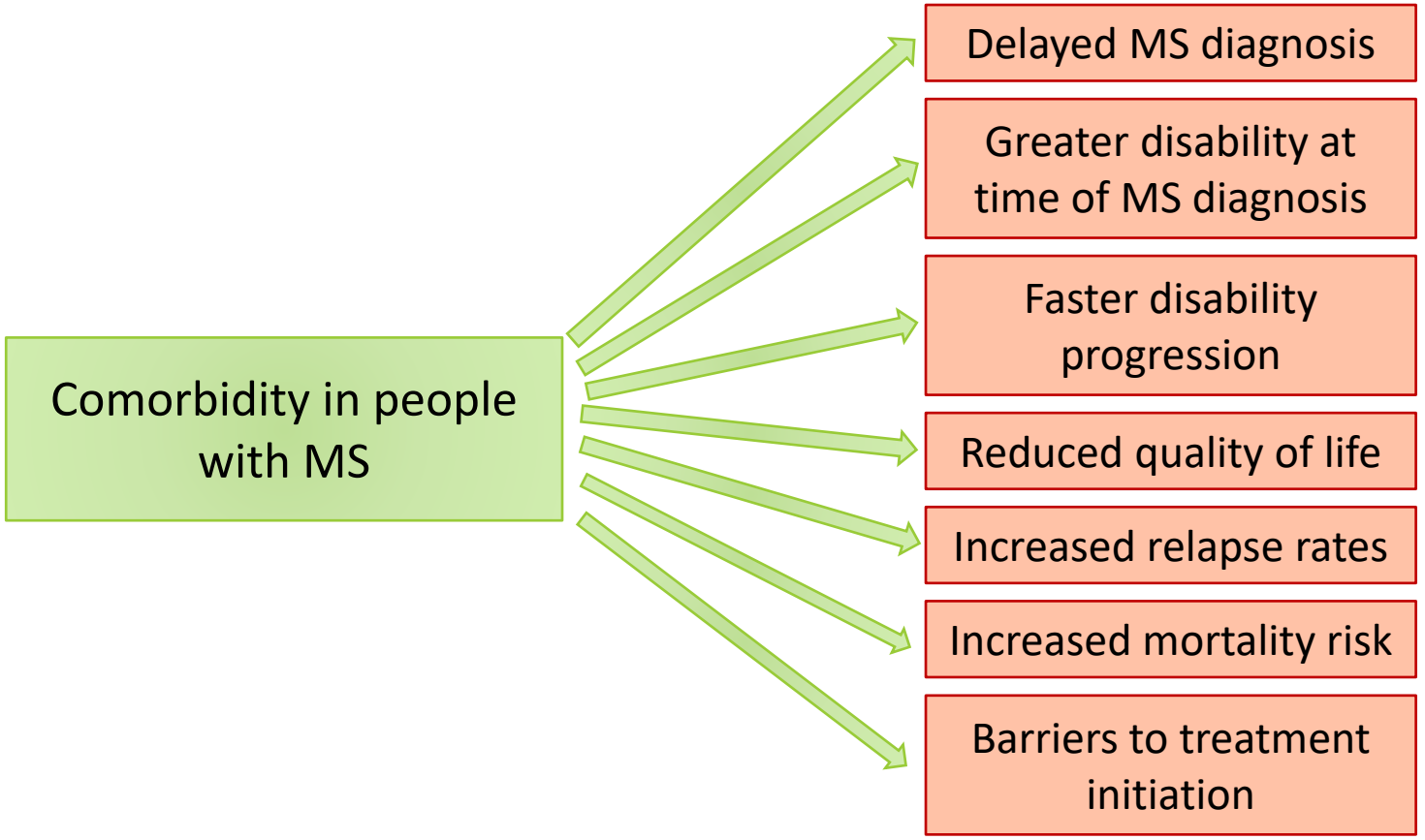
Assigned weights for each condition that a patient has. The total equals the score. Example: chronic pulmonary (1) and lymphoma (2) = total score (3).

Included conditions are comorbidities for the population of interest.

Charlson, ME. J Chronic Dis 1987;40:373-383.

$$\text{Score} = weight_1 * comorbidity_1 + weight_2 * comorbidity_2 + \cdots + weight_n * comorbidity_n$$

Why do we need a comorbidity index for people with MS?



Why do we need a comorbidity index for people with MS?

- Commonly used comorbidity summary measures have **not been validated** in MS populations

Measure	Population	Outcomes	Methodologic considerations
Charlson Comorbidity Index (1987)	Development: Patients at single New York **hospital** External: Breast cancer patients at New York **hospital**	**1-year mortality**	• **No internal validation** • Developed as an additive score
Elixhauser Comorbidity Index (1998)	Development: **Hospital** patients in California	• **In-hospital mortality** • **Length of stay in hospital** • Hospital charges	• **No external validation**

Why do we need a comorbidity index for people with MS?

❖ Commonly used comorbidity summary measures were **created for outcomes that may not be relevant** to the MS population

Measure	Population	Outcomes	Methodologic considerations
Charlson Comorbidity Index (1987)	Development: Patients at single New York **hospital** External: Breast cancer patients at New York **hospital**	**1-year mortality**	• **No internal validation** • Developed as an additive score
Elixhauser Comorbidity Index (1998)	Development: **Hospital** patients in California	• **In-hospital mortality** • **Length of stay in hospital** • Hospital charges	• **No external validation**

MS outcomes we will use in our index:

1. *Time to* disability progression
2. *Time to* treatment initiation
3. *Time to* death

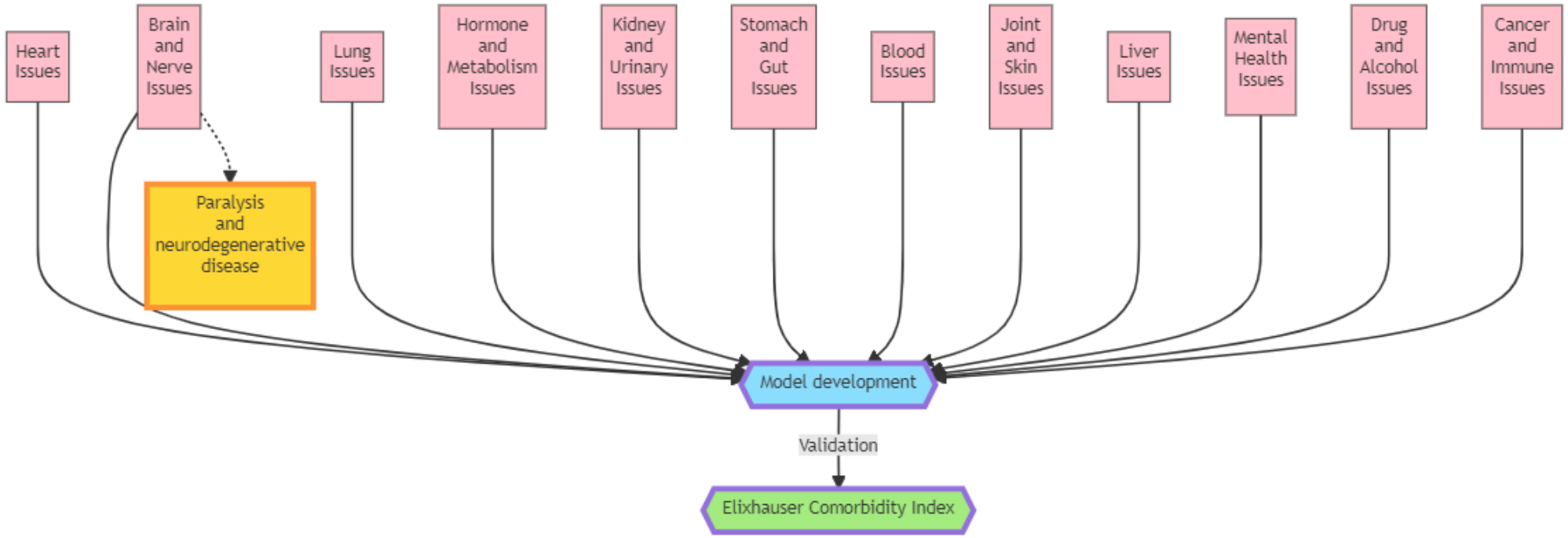
Why do we need a comorbidity index for people with MS?

- Commonly used comorbidity summary measures were created prior to the creation of **current prediction modeling guidelines**

Measure	Population	Outcomes	Methodologic considerations
Charlson Comorbidity Index (1987)	Development: Patients at single New York **hospital** External: Breast cancer patients at New York **hospital**	**1-year mortality**	• **No internal validation** • Developed as an additive score
Elixhauser Comorbidity Index (1998)	Development: **Hospital** patients in California	• **In-hospital mortality** • **Length of stay in hospital** • Hospital charges	• **No external validation**

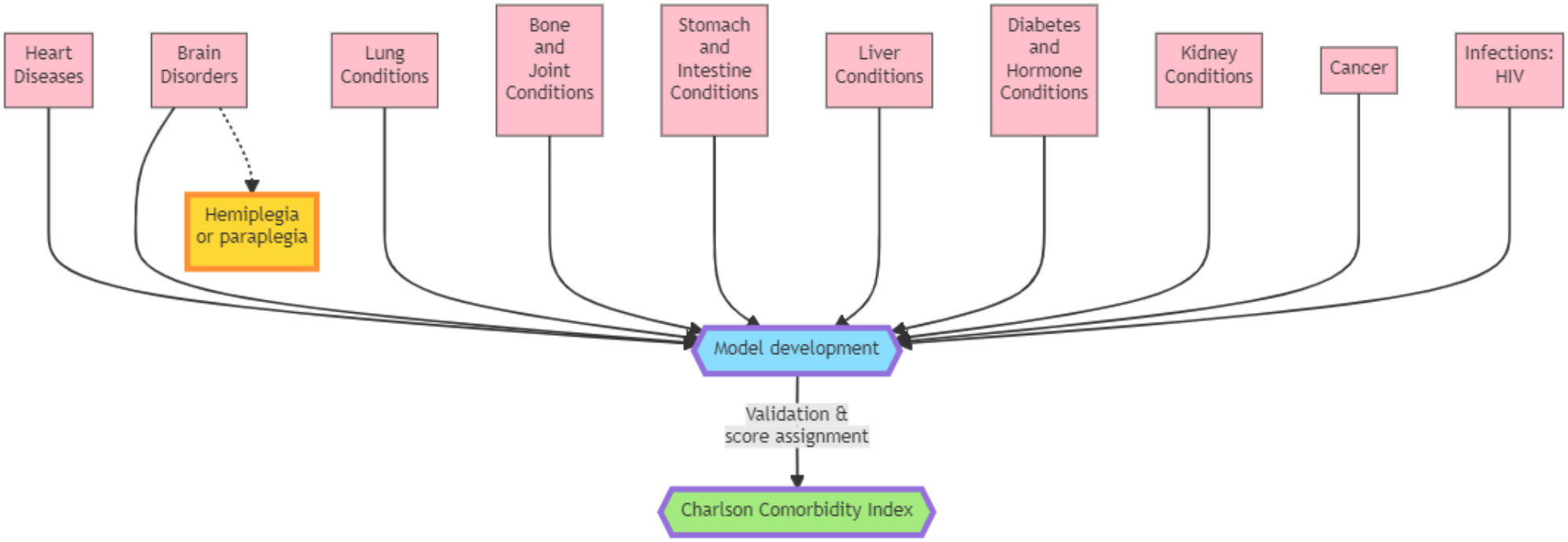
Why do we need a comorbidity index for people with MS?

- Commonly used comorbidity summary measures **include symptoms** of MS



Why do we need a comorbidity index for people with MS?

- Commonly used comorbidity summary measures **include symptoms** of MS



What comorbidities should we include?

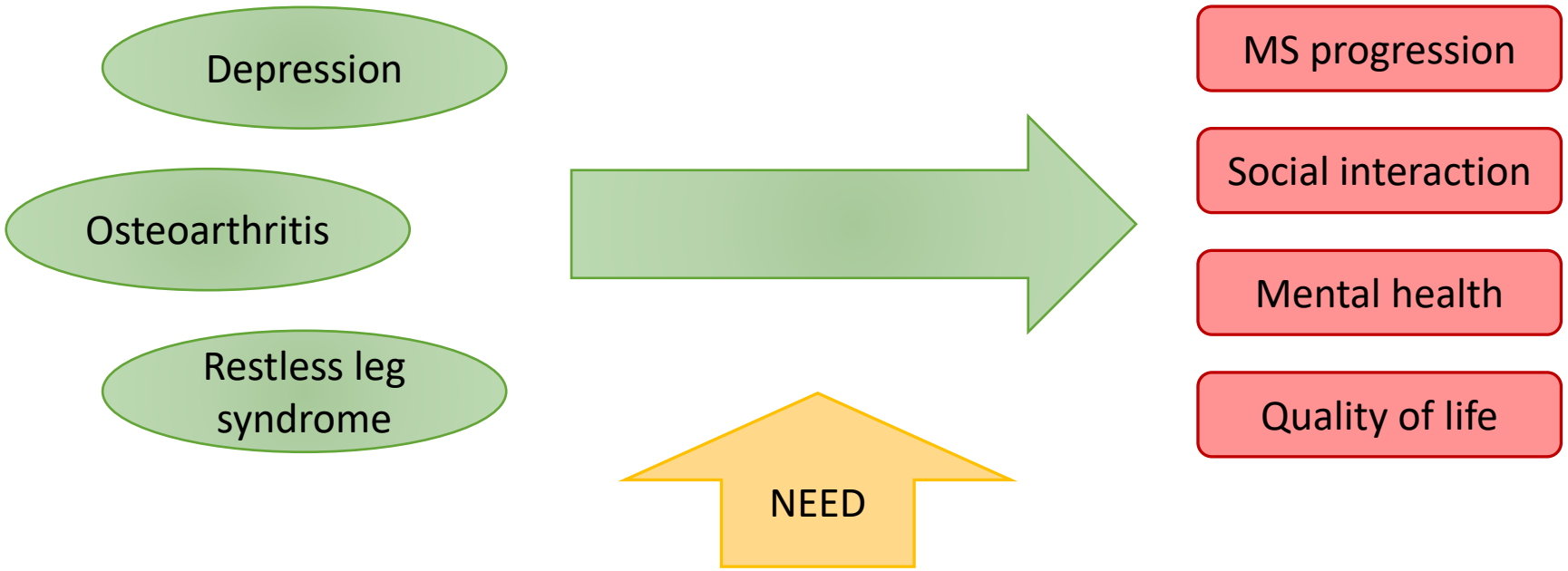
- Literature review

Type of condition	Substantial evidence of a potential association with…		
	Disability	Treatment initiation*	Mortality
Psychiatric	✓	✓	✓
Cardiovascular	✓	✓	?
Metabolic	✗	✓	✓
Autoimmune	✗	?	✗
Respiratory	✗	✓	?
Neurological	✓	✗	✓

* Treatment initiation was only examined by 2 studies

What comorbidities should we include?

- Discussions with people with lived experience & clinicians treating people with MS



Creating the index

- Cox model with comorbidities included based on a univariate analysis of associations
  - Performs **better than other measures**, but is **not good enough**

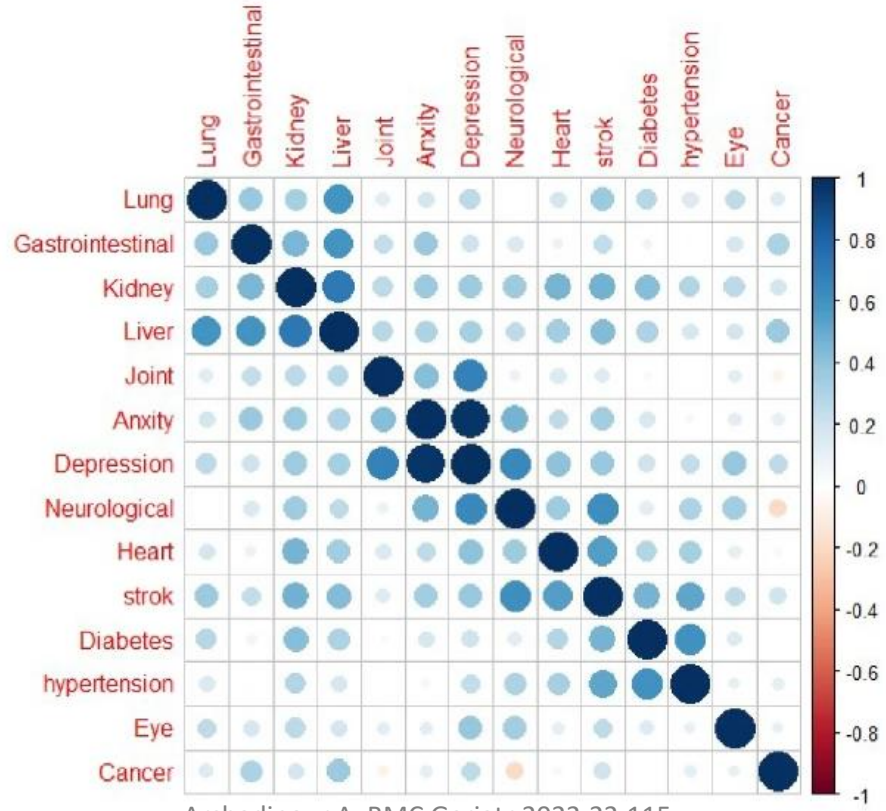
Index	C-statistic range	C-statistic interpretation*
1. Measure with all conditions found to be significant in log-rank tests[±]	0.61-0.70	Moderate discriminative power
2. Modified ECI with prevalent conditions[‡]	0.61-0.70	Moderate discriminative power
3. ECI	0.50-0.60	Low discriminative power
4. ECI excluding paralysis & neurological	0.50-0.60	Low discriminative power
5. ECI summary measure	0.50-0.60	Low discriminative power
6. Number of physician visits	0.50-0.60	Low discriminative power
7. ECI summary measure excluding paralysis & neurological	0.50-0.60	Low discriminative power
8. Number of diagnoses	0.50-0.60	Low discriminative power
9. CCI excluding hemiplegia	0.50-0.60	Low discriminative power
10. CCI	0.50-0.60	Low discriminative power

[±]In the British Columbia Multiple Sclerosis cohort : included depression, anxiety, psychoses, pulmonary heart disease, venous thromboembolism, stroke, heart failure, peripheral vascular disease, hyperlipidemia, thyroid disease, insomnia, glaucoma, myopia, ovarian cysts, endometriosis, female infertility, irritable bowel syndrome, gastroesophageal reflux disease, lung disease, hay fever, neoplasm of the colon, neoplasm of the skin, neoplasm of the uterus, mammary dysplasia, prostate hyperplasia, bladder cancer, epilepsy, migraines, renal disease, gout, ankylosing spondylitis, enthesopathy, synovitis, chronic skin ulcer, parkinsonism, dementia; [‡]Kang, JH. Eur Neurol 2010;17:1215-1219; *Syriha A. Cancer Med 2024;13:e6825.

Creating the index: How machine learning can help

- Health administrative data is **rich** - thousands of ICD-9 and -10 diagnosis codes exist
  - Many comorbidities are likely **correlated**:
    - Hypertension & diabetes
    - Depression & anxiety
    - Rheumatic conditions & depression

- Some machine learning methods can perform **variable selection** based on each variable's predictive power
  - **Least absolute shrinkage and selection operator (LASSO)**
    - Through **regularization** (limiting magnitude of coefficients), reduces some coefficients to 0



Arshadipour A. BMC Geriatr 2022;22:115.

Creating the index: How machine learning can help

- Interaction terms
  - Some comorbidities may have **interaction effects**: the effect of one comorbidity changes based on the presence of another comorbidity
  - Theoretical example:

    the effect of depression on disability may be greater if a person already has anxiety

- Machine learning methods can **identify and model** complex relationships/rules
  - Or find not yet well-understood covariate relationships
  - This reduces the risk of **model misspecification**
  - **Random survival forests**
    - **Automatically detects** interactions, higher-order terms
    - Can also perform **variable selection** based on variable importance measures

Arshadipour A. BMC Geriatr 2022;22:115.

How can we create a score from a machine learning model?

- **Option 1:** Rely on machine learning models to create a **set of comorbidities** similar to the Elixhauser index via variable selection
  - Each comorbidity in the set would be adjusted for individually in an analysis

- **Option 2:** Derive **weights** from the machine learning models
  - Not as easy as with a regression model
  - Regularization models (LASSO) set some coefficients to 0 – but this **may not be consistent** across iterations when using cross-validation
    - Need to decide when to exclude variables
  - Some machine learning models (random forests) **do not produce coefficients**
    - Could potentially use variable importance measures instead, but this has not yet been tested

Other important considerations: External validation

- External validation is critical for ensuring **generalizability**
- We will validate in:
  - **Manitoba**, a similar yet distinct Canadian population, and
  - **Sweden**, population with potentially different characteristics

Summary

- Comorbidity indices are important tools for observational research, but current indices need to be **updated** and **tailored** to specific diseases or outcomes
- Health administrative data is **rich** and used increasingly often for observational research
- Machine learning methods are useful because they can perform better than regression models when there is **multicollinearity** or **complex covariate relationships**
- One of the main **challenges** with using machine learning models for this task is that they do not always produce outputs comparable to regression models
  - This makes extracting weights difficult
- **Internal and external validation** steps are critical for ensuring we avoid overfitting and maximize the model's **generalizability**

Thank you. Questions?