(TO WHAT EXTENT) CAN REGISTRY DATA BE USED AS A RARE DISEASE DATA RESOURCE?

Jonathan Broomfield

✉ jb781@le.ac.uk

𝕏 @JonnyBroomfield

RARE DISEASE RESEARCH

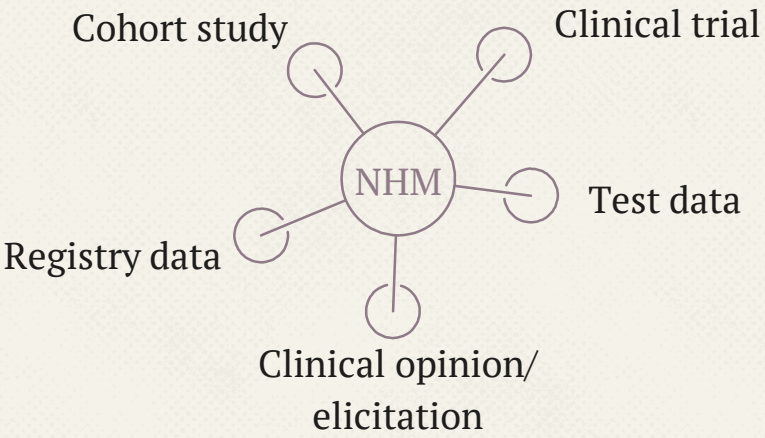
- Rare disease datasets are scarce…
- …small…
- …and heterogeneous

- Natural history models (NHMs) require data over the lifetime of people with a disease

RARE DISEASE RESEARCH

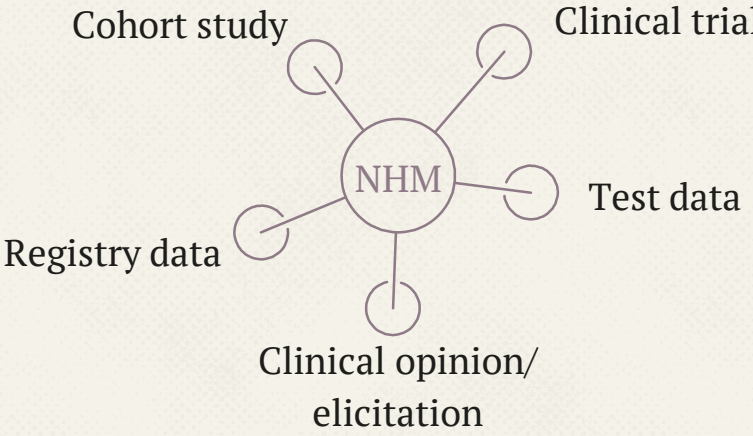
- How can we appropriately combine all available evidence into a rare disease NHM?

- How can we appropriately combine all available evidence into a rare disease NHM?



RARE DISEASE RESEARCH

- How can we appropriately combine all available evidence into a rare disease NHM?



Received: 13 March 2023 | Revised: 5 September 2023 | Accepted: 16 October 2023

DOI: 10.1002/sim.9949

RESEARCH ARTICLE

Statistics in Medicine WILEY

Modeling the multi-state natural history of rare diseases with heterogeneous individual patient data: A simulation study

Jonathan Broomfield[1] | Keith R. Abrams[2,3] | Suzanne Freeman[1] | Nicholas Latimer[4] | Mark J. Rutherford[1] | Michael J. Crowther[5] | On behalf of Project HERCULES, the Cooperative International Neuromuscular Research Group investigators and Duchenne Regulatory Science Consortium members

[1]Biostatistics Research Group, Department of Population Health Sciences, University of Leicester, Leicester, UK

[2]Department of Statistics, University of Warwick, Coventry, UK

[3]Centre for Health Economics, University of York, York, UK

[4]School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, UK

[5]Red Door Analytics, Stockholm, Sweden

Correspondence
Jonathan Broomfield, Biostatistics Research Group, Department of Population Health Sciences, University of Leicester, Leicester, UK.
Email: jb361@le.ac.uk

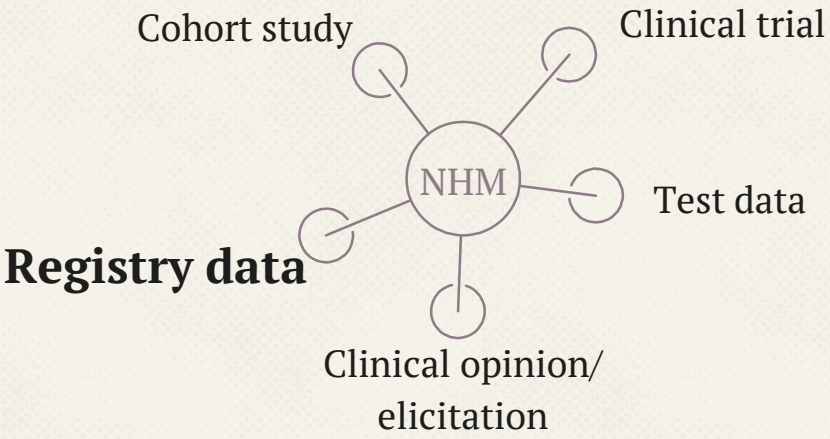
Funding information
National Institute for Health and Care Research, Grant/Award Number: NIHR300984

Multi-state survival models are used to represent the natural history of a disease, forming the basis of a health technology assessment comparing a novel treatment to current practice. Constructing such models for rare diseases is problematic, since evidence sources are typically much sparser and more heterogeneous. This simulation study investigated different one-stage and two-stage approaches to meta-analyzing individual patient data (IPD) in a multi-state survival setting when the number and size of studies being meta-analyzed are small. The objective was to assess methods of different complexity to see when they are accurate, when they are inaccurate and when they struggle to converge due to the sparsity of data. Biologically plausible multi-state IPD were simulated from study- and transition-specific hazard functions. One-stage frailty and two-stage stratified models were estimated, and compared to a base case model that did not account for study heterogeneity. Convergence and the bias/coverage of population-level transition probabilities to, and lengths of stay in, each state were used to assess model performance. A real-world application to Duchenne Muscular Dystrophy, a neuromuscular rare disease, was conducted, and a software demonstration is provided. Models not accounting for study heterogeneity were consistently out-performed by two-stage models. Frailty models struggled to converge, particularly in scenarios of low heterogeneity, and predictions from models that did converge were also subject to bias. Stratified models may be better suited to meta-analyzing disparate sources of IPD in rare disease natural history/economic modeling, as they converge more consistently and produce less biased predictions of lengths of stay.

KEYWORDS
multi-state model, natural history, rare diseases, simulation, survival analysis

[1]

- How can we appropriately combine all available evidence into a rare disease NHM?



Received: 13 March 2023 | Revised: 5 September 2023 | Accepted: 16 October 2023
DOI: 10.1002/sim.9949

RESEARCH ARTICLE

Statistics in Medicine WILEY

Modeling the multi-state natural history of rare diseases with heterogeneous individual patient data: A simulation study

Jonathan Broomfield[1] | Keith R. Abrams[2,3] | Suzanne Freeman[1] | Nicholas Latimer[4] | Mark J. Rutherford[1] | Michael J. Crowther[5] | On behalf of Project HERCULES, the Cooperative International Neuromuscular Research Group investigators and Duchenne Regulatory Science Consortium members

[1]Biostatistics Research Group, Department of Population Health Sciences, University of Leicester, Leicester, UK
[2]Department of Statistics, University of Warwick, Coventry, UK
[3]Centre for Health Economics, University of York, York, UK
[4]School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, UK
[5]Red Door Analytics, Stockholm, Sweden

Correspondence
Jonathan Broomfield, Biostatistics Research Group, Department of Population Health Sciences, University of Leicester, Leicester, UK.
Email: jb361@le.ac.uk

Funding information
National Institute for Health and Care Research, Grant/Award Number: NIHR300984

Multi-state survival models are used to represent the natural history of a disease, forming the basis of a health technology assessment comparing a novel treatment to current practice. Constructing such models for rare diseases is problematic, since evidence sources are typically much sparser and more heterogeneous. This simulation study investigated different one-stage and two-stage approaches to meta-analyzing individual patient data (IPD) in a multi-state survival setting when the number and size of studies being meta-analyzed are small. The objective was to assess methods of different complexity to see when they are accurate, when they are inaccurate and when they struggle to converge due to the sparsity of data. Biologically plausible multi-state IPD were simulated from study- and transition-specific hazard functions. One-stage frailty and two-stage stratified models were estimated, and compared to a base case model that did not account for study heterogeneity. Convergence and the bias/coverage of population-level transition probabilities to, and lengths of stay in, each state were used to assess model performance. A real-world application to Duchenne Muscular Dystrophy, a neuromuscular rare disease, was conducted, and a software demonstration is provided. Models not accounting for study heterogeneity were consistently out-performed by two-stage models. Frailty models struggled to converge, particularly in scenarios of low heterogeneity, and predictions from models that did converge were also subject to bias. Stratified models may be better suited to meta-analyzing disparate sources of IPD in rare disease natural history/economic modeling, as they converge more consistently and produce less biased predictions of lengths of stay.

KEYWORDS
multi-state model, natural history, rare diseases, simulation, survival analysis

[1]

Clinical Practice Research Datalink

- *UK-based electronic health record*
- *GP data that can be linked to HES, ONS…*

A CASE STUDY IN DUCHENNE MUSCULAR DYSTROPHY (DMD)

Rare, muscle-wasting disease

Global prevalence of 1 in 3,500 live male births [2]

Currently there is no cure for DMD

Steroid use
Spinal surgery
Ventilation
Cardiomyopathy
Mortality

EXTRACTING EVENTS OF INTEREST

Steroid use

- Defined product code lists for oral corticosteroids
- Took first steroid use as event of interest

Spinal surgery

- Identified using HES records with OPCS codes V22*-V70*
- Age taken as midpoint of HES episode

Ventilation

- Identified using HES records with OPCS codes E85*
- Age taken as midpoint of HES episode

Cardiomyopathy

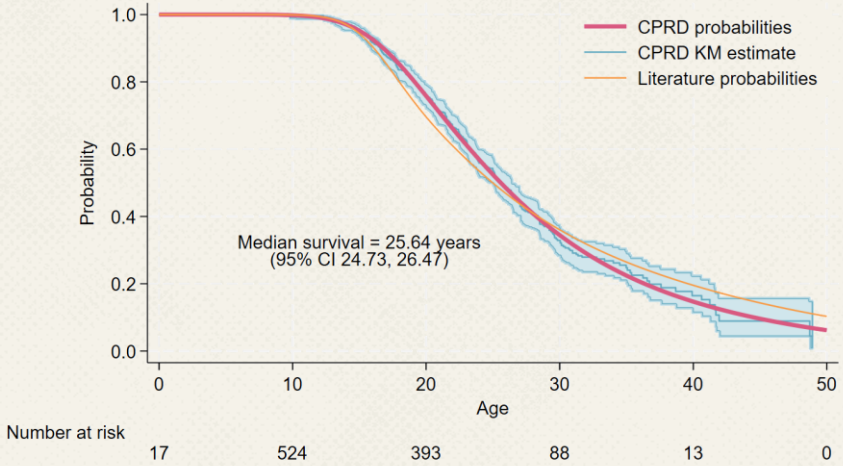
- Defined medical code lists for cardiomyopathy
- Inspected ONS cause of death records in secondary analysis

Mortality

- Taken from ONS records, or CPRD records if missing from ONS
- Latest follow-up time from CPRD/ HES used as censoring time

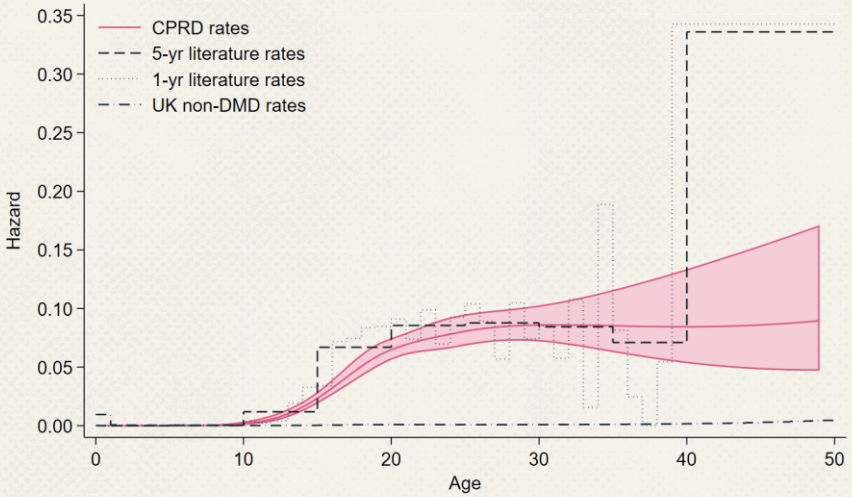
MORTALITY

[3]



KM plot of survival in the CPRD cohort with overlain predictions from a flexible parametric model, left-truncated at practice registration dates, and estimates from a review of international studies [4].

[3]



Mortality rates in the CPRD cohort from a flexible parametric model, left-truncated at practice registration dates, overlain with estimates from a review of international studies [4] and from the UK general population.

INTERMEDIATE EVENTS

Summary statistics of ages at clinical milestones of DMD. [3]

Milestone	# events (%)	Median (95% CI)	Mean (95% CI)
First corticosteroid	479	6.06	6.34
	(46.7)	*(5.77, 6.29)*	*(6.11, 6.56)*
Spinal surgery	155	14.79	14.68
	(15.0)	*(14.29, 15.09)*	*(14.34, 15.01)*
Ventilation	286	16.97	18.12
	(27.1)	*(16.50, 18.31)*	*(17.38, 18.87)*
Cardiomyopathy	133	15.26	15.17
	(12.5)	*(14.22, 16.70)*	*(14.08, 16.27)*

Validity

We do not observe patients across their lifetimes

Generalisability

Is the CPRD representative of all UK patients?

Reproducibility

Can we update the model over time?

REFERENCES

1. Broomfield J, Abrams K, Freeman S, Latimer M, Rutherford M, Crowther M. (2024) Modelling the multi-state natural history of rare diseases with heterogeneous individual patient data: a simulation study. *Statistics in Medicine*, **43**(1), 184-200.

2. National Organisation for Rare Diseases (NORD). (2024) *Duchenne Muscular Dystrophy*. Available at: https://rarediseases.org/rare-diseases/duchenne-muscular-dystrophy/ [Accessed: 23/02/2024].

3. Broomfield J, Abrams K, Latimer M, Guglieri M, Rutherford M, Crowther M. (2023) Natural history of Duchenne Muscular Dystrophy in the UK: A descriptive study using the Clinical Practice Research Datalink. *Brain and Behavior*, **13**, e3331.

4. Broomfield J, Hill M, Guglieri M, Crowther M, Abrams K. (2021). Life expectancy in Duchenne Muscular Dystrophy: reproduced individual patient data meta-analysis. *Neurology*, **97**(23), e2304-e2314.