Keynote presentation at the Academy of Medical Sciences Joint Workshop
Advancing research to tackle multimorbidity: the UK perspective

Methodological approaches to multimorbidity research

Sylvia Richardson

MRC Biostatistics Unit

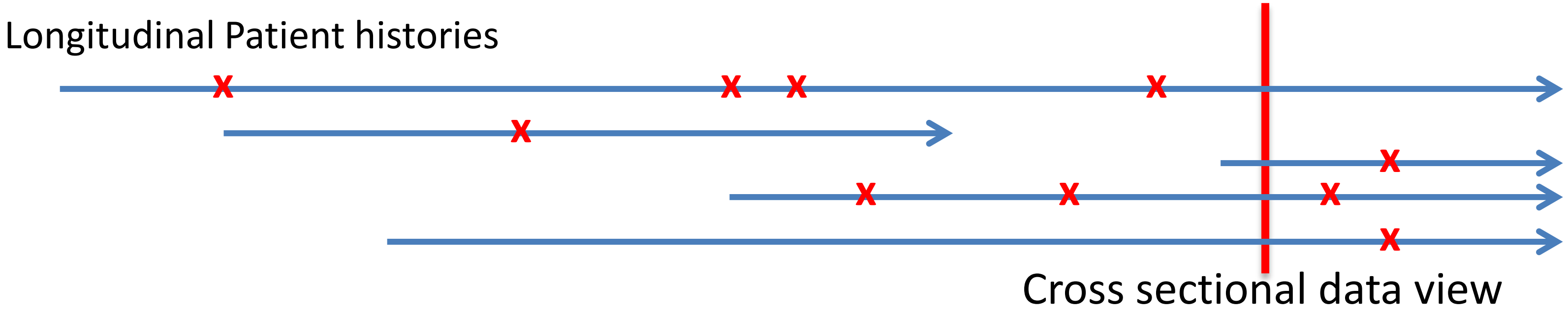
Thanks to Jessica Barrett, Steven Kiddle, Kirsty Rhodes, Li Su & Brian Tom
(MRC Biostatistics Unit)
Simon Griffin (MRC Epidemiology Unit)
and Emanuele Di Angelantonio (Cardiovascular Epidemiology Unit)
for fruitful discussions

Research priorities

- The main key words of the outlined six research priorities interface with different methodological approaches and needs:
  - Trends and patterns: descriptive → data quality
  - Clusters of conditions: descriptive → data compression
  - Burden: analytic → estimation of link, prediction of risk
  - Determinants: analytic → event history modelling
  - Benefits and risks of treatment: beyond RCT → trial emulation to account for multimorbidity.

Research priorities

- The main key words of the outlined six research priorities interface with different methodological approaches and needs:
  - Trends and patterns: descriptive → data quality
  - Clusters of conditions: descriptive → data compression
  - Burden: analytic → estimation of link, prediction of risk
  - Determinants: analytic → event history modelling
  - Benefits and risks of treatment: beyond RCT → trial emulation to account for multimorbidity.



1 Data issues

- Central issue of adopting a coherent definition of comorbidity has been highlighted in the report.
- Much of the descriptive research is cross-sectional and utilises large primary care databases.
- A recent Danish study used their national patient registry of all hospitals.
- UK studies have relied on GPs recording conditions, by giving patients diagnosis codes or by prescribing medications.

Issues that have been discussed

- inconsistent labelling of conditions, choice of granularity,
- Under reporting, differential validity in assessment of conditions,
- How representative are the sample of GP practices that are used (e.g. CPRD)?

1 Data issues

- Central issue of adopting a coherent definition of comorbidity has been highlighted in the report.
- Much of the descriptive research is cross-sectional and utilises large primary care databases.
- A recent Danish study used their national patient registry of all hospitals.
- UK studies have relied on GPs recording conditions, by giving patients diagnostic codes or by prescribing medications.

Issues that have been discussed:

- inconsistent labelling of conditions, choice of granularity,
- Under reporting, differential validity in assessment of conditions,
- How representative are the sample of GP practices that are used (e.g. CPRD)?

- Careful assessment of data quality, sources of missingness, mismeasurement and biases is of crucial importance alongside any pattern extraction.

  → Some of these issues might be addressed through specially designed calibration studies.

Recent cross sectional study investigating multimorbidity in CPRD

Patient	Hypertension	Depression	Diabetes	Asthma	…	Cancer
1	0	1	0	0	…	0
2	1	0	0	0	…	1
3	1	1	0	1	…	0
4	1	0	0	0	…	0
5	1	0	1	0	…	0
…	…	…	…	…	…	…
403,985	1	1	0	0	…	0

Binary comorbidity matrix X

Dataset

- 403,985 adult patients
- 37 long-term conditions coded
- Recorded characteristics:
  - Gender
  - Age
  - Socioeconomic status
- Health service utilisation outcomes:
  - General practice consultations
  - Prescriptions
  - Hospitalisations

Cassell et al. (2018), British Journal of General Practice

Recent cross sectional study investigating multimorbidity in CPRD

Table 2. Ten most prevalent morbidities and associated comorbidities

Morbidity	Prevalence, %	Mean number of comorbidities associated with condition, n	Three most frequently associated comorbidities: Condition	Prevalence,[a] %
Hypertension	18.2	3.0	Painful condition	24.3
			Diabetes	19.4
			Hearing loss	16.7
Depression/anxiety	10.3	3.1	Painful condition	32.7
			Hypertension	28.9
			Irritable bowel syndrome	17.2
Chronic pain	10.1	3.7	Hypertension	44.0
			Depression/anxiety	35.5
			Hearing loss	18.4
Hearing loss	9.5	2.8	Hypertension	32.0
			Painful condition	19.4
			Depression/anxiety	14.8
Irritable bowel syndrome	7.9	1.8	Depression/anxiety	22.3
			Hypertension	20.5
			Painful condition	18.4
Diabetes	5.9	3.5	Hypertension	60.1
			Painful condition	26.6
			Depression/anxiety	17.9
Prostate disorders	5.7	3.5	Hypertension	44.1
			Hearing loss	25.3
			Painful condition	20.7
Thyroid disorders	4.7	3.1	Hypertension	37.0
			Painful condition	23.4
			Depression/anxiety	19.7
Coronary heart disease	4.3	4.0	Hypertension	56.5
			Painful condition	30.3
			Diabetes	23.3
Asthma	3.7	3.2	Hypertension	30.3
			Painful condition	26.6
			Depression/anxiety	22.4

[a]Prevalence in male participants only.

Dataset

- 403,985 adult patients
- 37 long-term conditions coded

(http://www.phpc.cam.ac.uk/pcu/cprd_cam/codelists/)

- 27% of patients had multimorbidity
- Patients with multimorbidity accounted for 53% of GP consultations, 79% of prescriptions and 56% of hospitalisations.

Cassell et al. (2018), British Journal of General Practice

2 Clustering and finding patterns

- A number of unsupervised exploratory approaches have been implemented to extract patterns from the binary matrix X:

(cf Ng, 2014, Roso-Llorach, 2018) :

  - Hierarchical Clustering,
  - Exploratory Factor Analysis,
  - Bi-clustering.
- Typically these exploratory approaches rely on a sequence of processing steps
  - Many ad hoc choices in terms of analysis strategy, choice of similarity metric, number of clusters, etc.
  - Problems of interpretation.

Patient	Hypertension	Depression	Diabetes	Asthma	…	Cancer
1	0	1	0	0	…	0
2	1	0	0	0	…	1
3	1	1	0	1	…	0
4	1	0	0	0	…	0
5	1	0	1	0	…	0
…	…	…	…	…	…	…
n	1	1	0	0	…	0

2 Clustering and finding patterns

- A number of unsupervised exploratory approaches have been implemented to extract patterns from the binary matrix X:

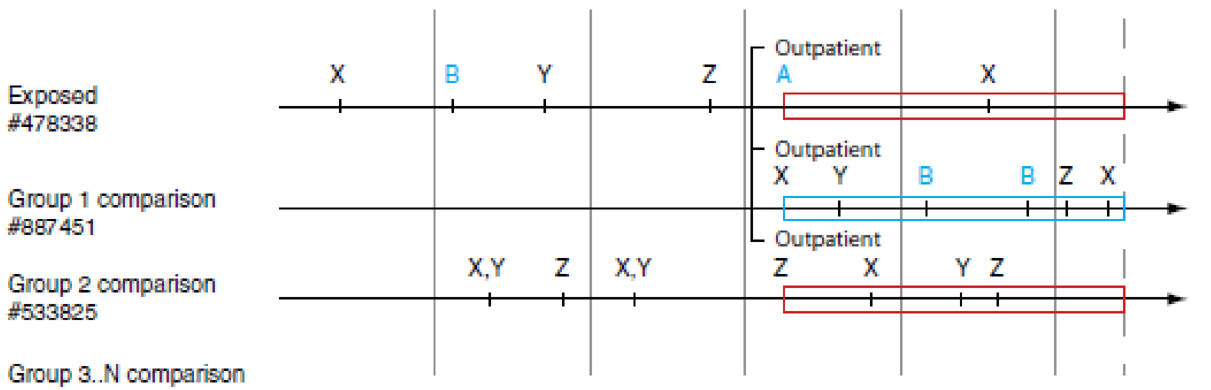
(cf Ng, 2014, Roso-Llorach, 2018) :
  - Hierarchical Clustering,
  - Exploratory Factor Analysis,
  - Bi-clustering.
- Typically these exploratory approaches rely on a sequence of processing steps:
  - Many ad hoc choices in terms of analysis strategy, choice of similarity metric, number of clusters, etc.
  - Problems of interpretation.

Patient	Hypertension	Depression	Diabetes	Asthma	…	Cancer
1	0	1	0	0	…	0
2	1	0	0	0	…	1
3	1	1	0	1	…	0
4	1	0	0	0	…	0
5	1	0	1	0	…	0
…	…	…	…	…	…	…
n	1	1	0	0	…	0

Would be useful to take a probabilistic perspective on how to learn the decomposition of binary matrix X into **two low dimensional matrices** corresponding to latent factors/ patterns and allocations.

Extracting patterns from longitudinal trajectories

- Longitudinal trajectories carry additional information
  - Investigate significant occurrences of pairs of diseases (A and B) and their temporality A → B, within a specified time frame.



from Jensen et al 2014

Extracting patterns from longitudinal trajectories

- Longitudinal trajectories carry additional information:
  - Investigate significant occurrences of pairs of diseases (A and B) and their temporality A → B, within a specified time frame.
  - Investigate significant occurrences of A → B → C, within a specified time frame.
  - Quickly becomes unwieldy "combinatorially"

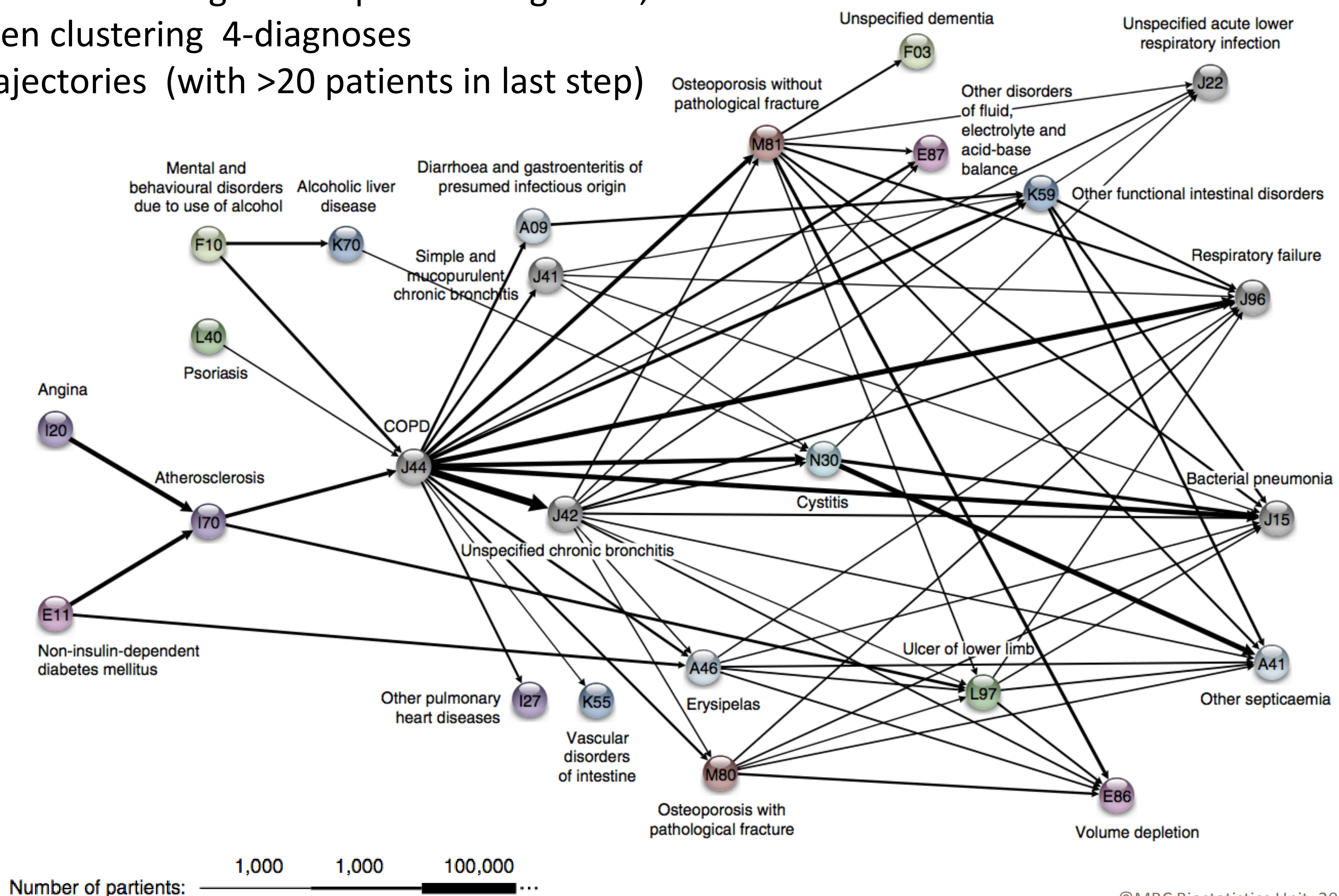
    → cluster the trajectories directly?
  - Main issue: defining relevant "distance function" between the trajectories to tease out meaningful patterns and groupings.

Specific context and aims should inform the choice of distance.

Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients – Jensen et al., (2014)

Methodology based on piecing together time ordered significant pairs of diagnoses, then clustering 4-diagnoses trajectories (with >20 patients in last step)

COPD disease `trajectory cluster'



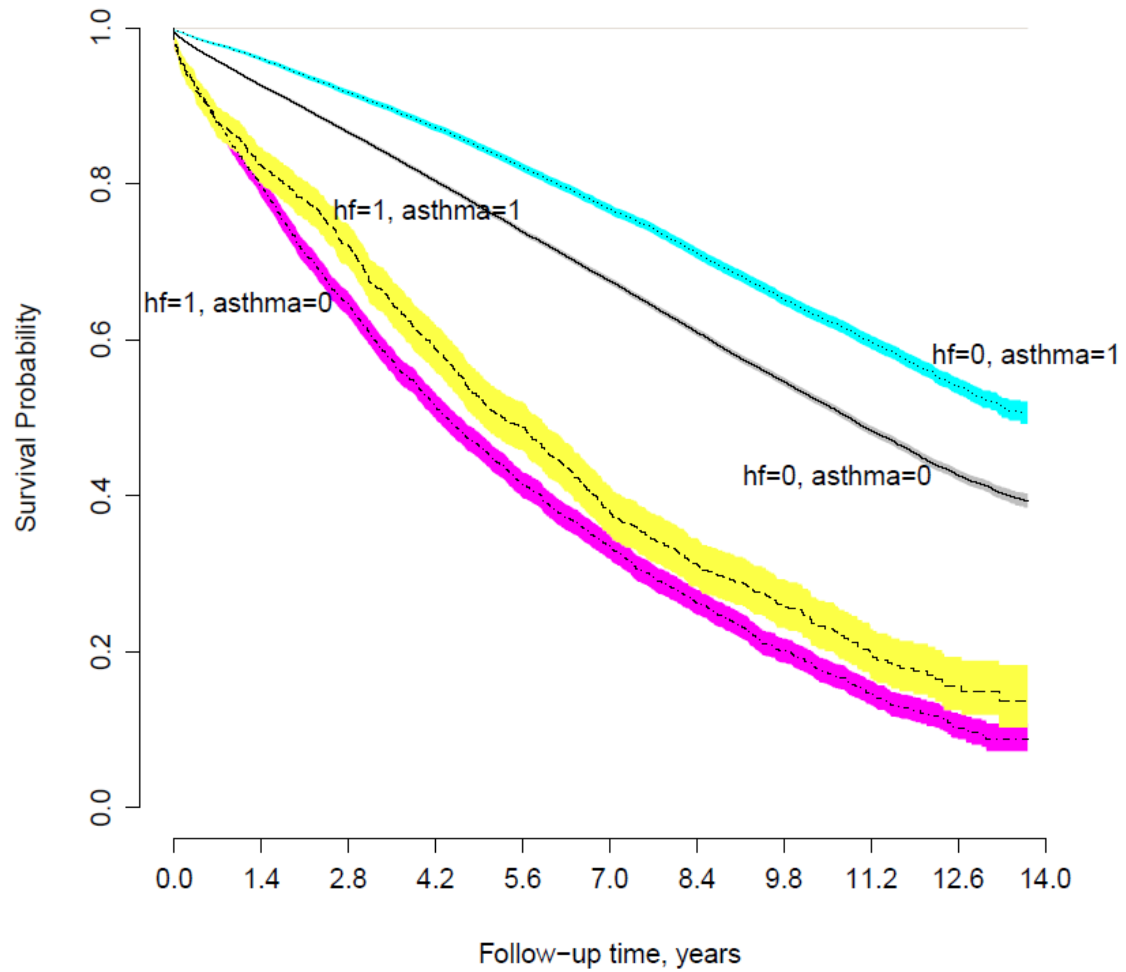
3 Burden

- Charlson co-morbidity Index as a summary measure of mortality risk is still popular.
  - Drawbacks
    - Additive formulation does not consider possibility of interactions between conditions,
    - Doesn’t consider treatment,
    - Doesn’t take into account patient history.

3 Burden

- Charlson co-morbidity Index as a summary measure of mortality risk is still popular.
  - Drawbacks
    - Additive formulation does not consider possibility of interactions between conditions,
    - Doesn't consider treatment
    - Doesn't take into account patient history.
- With longitudinal data, can use **survival models** to estimating link between multimorbidity and outcomes, such as death or disability.
  - Benefits: loss to follow-up is properly accounted for.
  - Survival models formulation can also be adapted to answer a number of focussed questions.
    - If the focus is on a number of index conditions, can estimate in turn the risk of developing another condition using other comorbidities as time dependent covariates.

Survival (Weibull model) after first diagnosis of COPD, stratified by history of asthma and heart failure



Study performed by Steven Kiddle (BSU), Hannah Whittaker & Jennifer Quint (Imperial College)

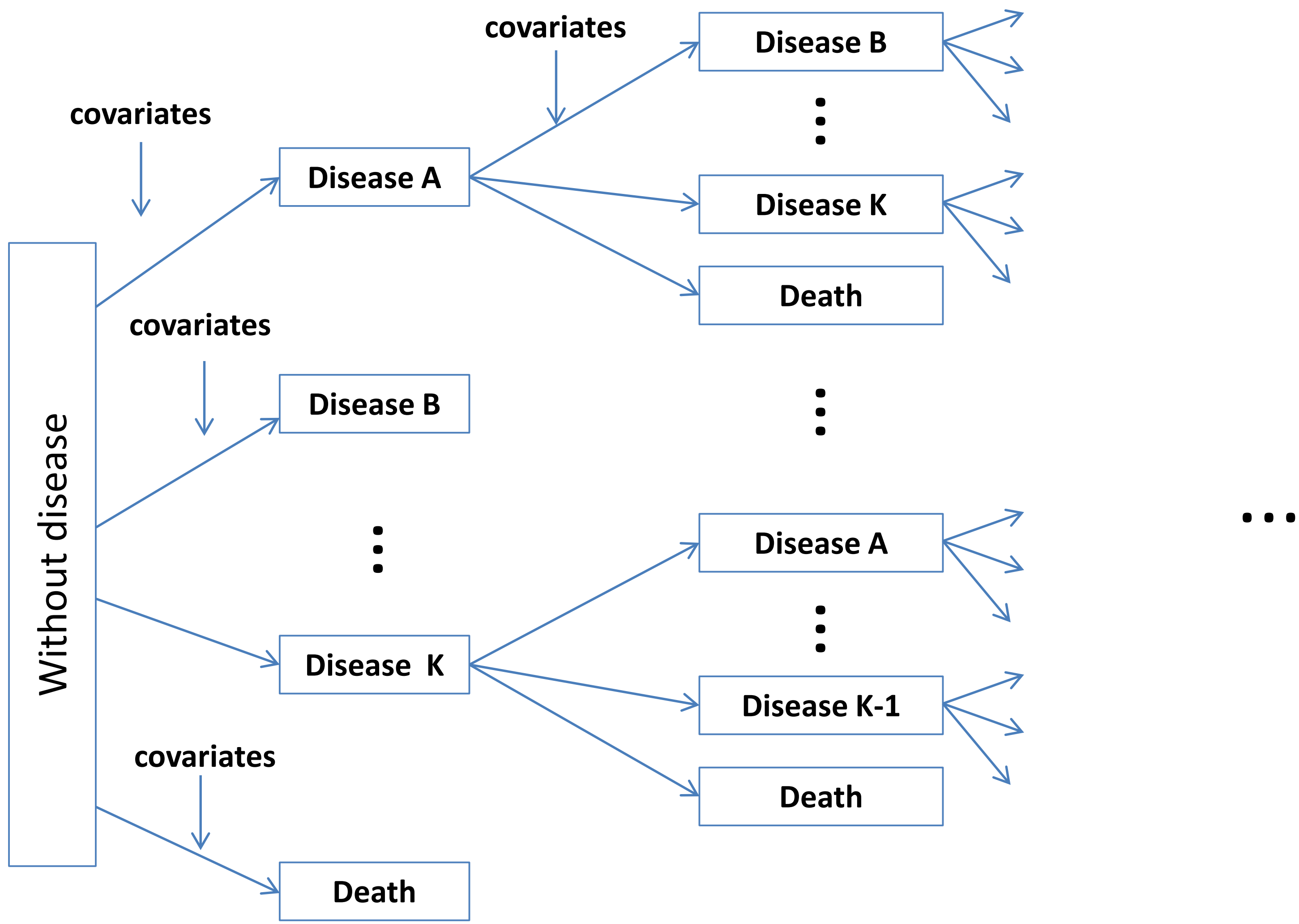
Burden

- Opportunity to use state of the art statistical modelling approaches coupled with survival models
  - to develop flexible multivariate survival models going beyond additive models and proportional hazard assumptions,
  - allowing possibility of interactions between the conditions,
  - taking into account patient history,
  - large model space to explore:
    - can benefit from advances in high dimensional regression approaches
    - model choice and validation through out-of-sample performance evaluation.

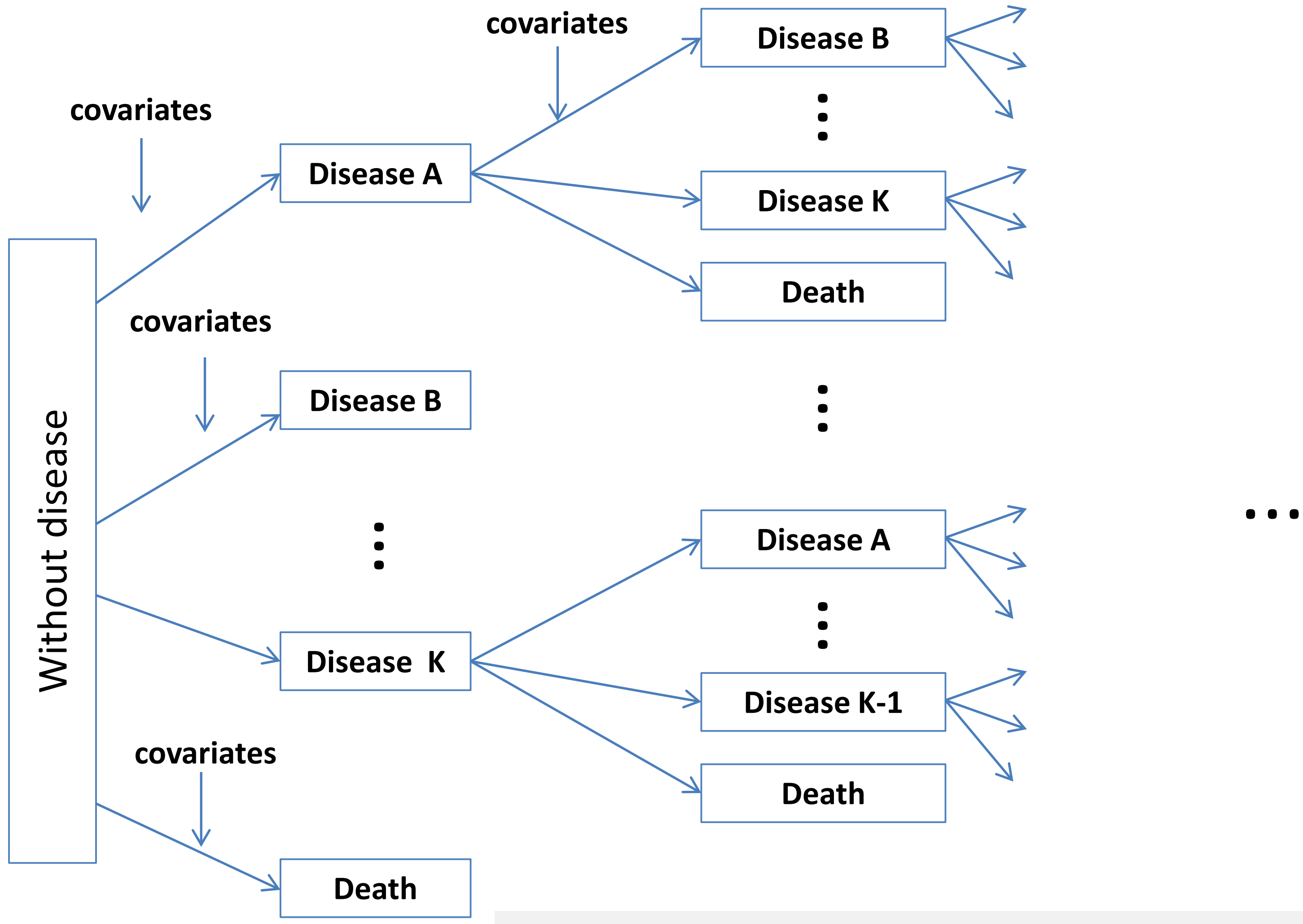
4 Determinants

- Best investigated using event history analysis tools, considering full history.
- In particular, multi-state models with estimation of transition rates and covariate effects on these transitions is a framework which would provide interpretable quantities.

Generic multi state model
of multimorbidity in patient trajectories



Generic multi state model
of multimorbidity in patient trajectories



Issues of computational scalability of multistate model estimation on large data bases

Determinants

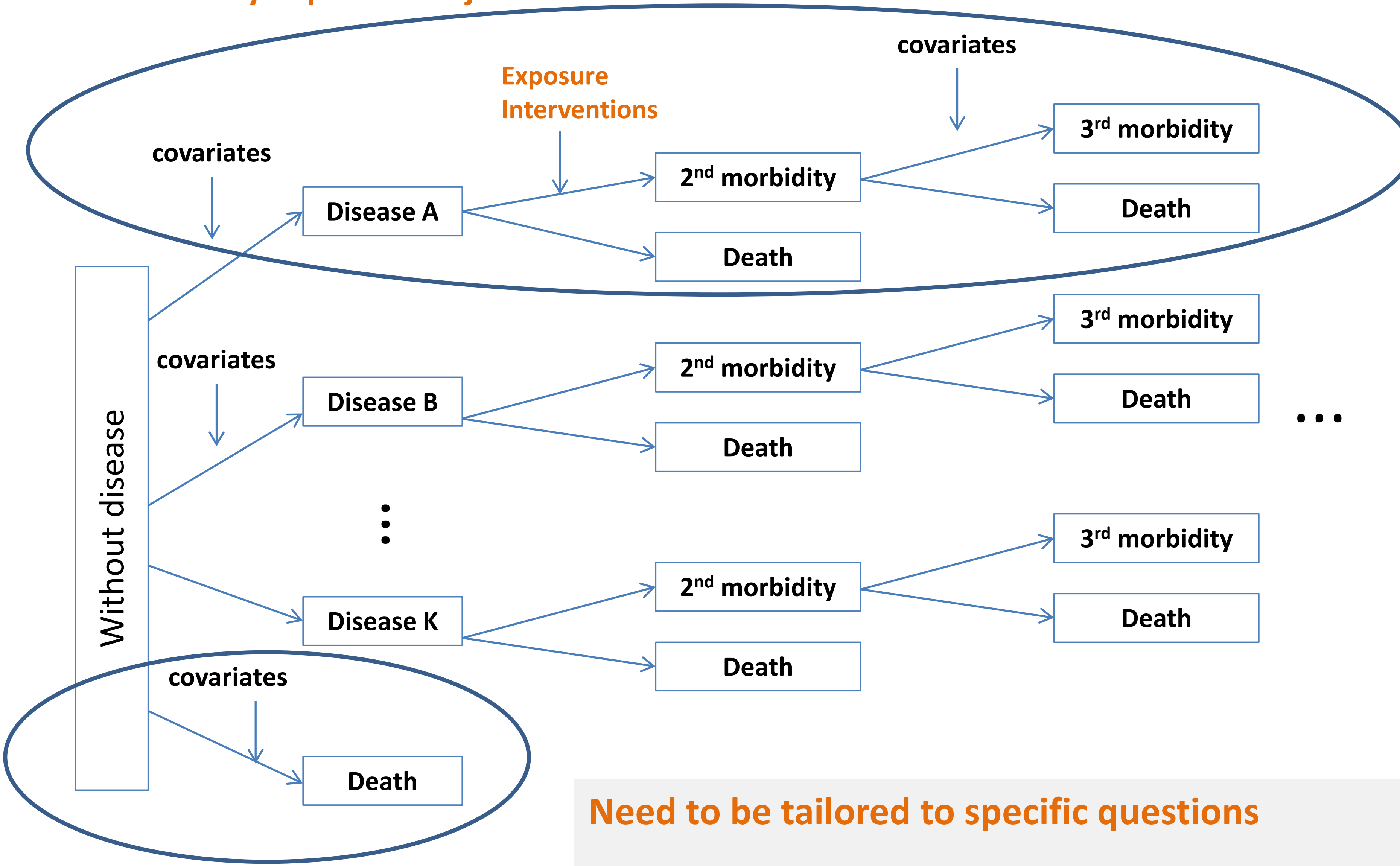
- Best investigated using event history analysis tools, considering full history.
- In particular, multi-state models with estimation of transition rates and covariate effects on these transitions is a framework which would provide interpretable quantities.

- Important issue: defining meaningful “states” and allowable transitions between these.

→ this has to be linked closely to epidemiological and medical context.

- Potential for formulating causal hypotheses to be investigated further.

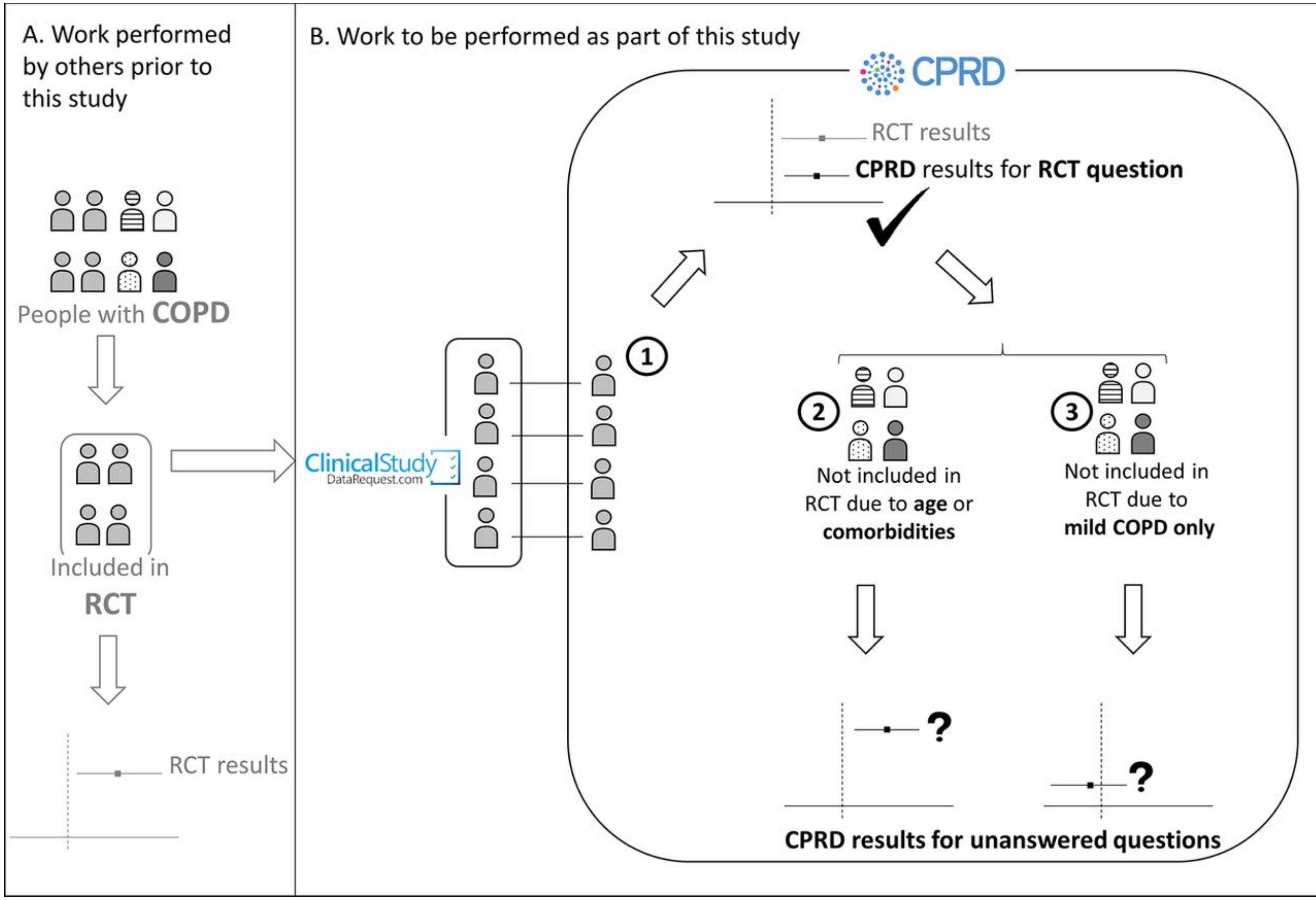
Simplified multi state model of multimorbidity in patient trajectories



Need to be tailored to specific questions

5 Benefits and risk of treatment in patients with multimorbidity

- Most RCTs exclude patients with multimorbidity
- Investigate trial emulation in large EHR data bases as a way to include a more realistic population of patients and address effect of treatment



Wing et al., (2018) BMJ open

In summary

- Rich range of methodological approaches can be tailored to characterise patterns of comorbidities, and to estimate burden and disease processes.
  - Important to take into account limitations of the data,
  - Important to confront generic statistical approaches with specific research questions to derive coherent analysis strategies.
- Longitudinal analysis is key to dig deeper into processes and determinants.
- Need to prioritize important health questions related to multimorbidity that can be reasonably tackled with large electronic data bases.