Trajectory Analysis

Megan Farris, MSc Epidemiologist

Medlior Health Outcomes Research Ltd; University of Calgary

R Calgary: October 17, 2018

Repeated measures studies

- Cohort studies, that have several assessments of:
  - The exposure
  - The outcome
  - Other potential covariates
- Very expensive to conduct
- Will provide a wealth of information in regards to causal mechanisms
- May account for change over time, prognosis, trajectories etc.

Background: Trajectory modeling

- Technique used to describe growth, prognosis, development over age or time
- May provide insights to disease etiology, the developmental course of different diseases and causal inference in epidemiological studies
- Fairly new technique to medicine but has also been applied to criminology and psychology fields
- Extension of Growth Curve Modeling (GMM)

Growth Curve Modeling (Does One Type of Curve Fit All?)

- Measure average development over time of a population (your study sample)
- Capture mean trends in development and individual departures from the average trend
- Assumptions: Your sample is drawn from the same population and classified under one trajectory curve
  - Sufficient for inter-individual variability
- **Implications?**
  - Biases?
- Sufficient for the study of general trends
  - e.g., cancer incidence trends over time
  - Economic burden of cancer screening over time

Growth Mixture Modeling

- Extension of GCM using finite mixture models (aka latent class models)
  - Provide a natural representation of heterogeneity in a finite number of latent classes (classes meaning groups/clusters of people)
  - Allows for variation of different distributions rather than just one distribution that fits all
- You need to distinguish the sub-populations based on actual measured characteristics
  - E.g., if you have people in your cohort with a sub-population with a genetic vulnerability and a sub-population without genetic vulnerabilities
  - You would analyze these groups separately as they might have different growth curves (depending on your research question).
- Implications?
- Would we always know what these sub-populations might be?

Group-Based Trajectory Modeling (GBTM)

- Based on finite mixture modeling as well
- Assume the population is made up of distinct groups defined by their development trajectories (driven by the data)
  - This phenomenon may not be physiologically or biologically correct, but that is what you are testing
  - The bigger motivation for these models is to draw attention to different characteristics or consequences of different trajectory groups
  - Research questions might look at: Are etiologic considerations and trajectory groups modeled actually present in the population? If so, what characteristics define those groups?
- Each trajectory group is thought of as a group of individuals who follow a similar development

Main difference between GMM and GBTM

- GMM assumes there are sub-populations follow a specific growth curve

- GBTM makes no population assumptions and uses the trajectory groups produced by statistical analysis to approximate unknown distribution of trajectories that might be present in the population

  - The theory behind this implies that statistical methods are sensitive to these differences in data

  - We then as Epidemiologists need to determine if these unknown population distributions are in actual fact, real clinically relevant sub-populations

Considerations of GBTM

- First, the number of groups and their functional form
  - How big is your sample?
  - How much variability can you expect within your sample?
  - *A priori* pick a maximum number of groups to be tested on your data
  - Fit number of groups based on several criteria
  - These include both statistical considerations as well as practical (i.e., actual differences between groups)

- Justify all decisions!

GBTM: Adding predictors

- Not only can you identify sub-populations within your study, you can characterize them (adding predictors to group membership)
- What characteristics are relevant to your study population, or that might classify individuals?
- Commonly collected a baseline or at study enrollment
- Besides descriptively, you can also use a multinomial logit model to compare characteristics of one group relative to another

You can take into account attrition

- It is common to have dropout in a prospective cohort study

- With GBTM you can:
  - Model the dropout into the estimation of trajectory groups
  - Test if dropout influences trajectory group assignment
  - Determine if there is differential dropout between trajectory groups

Add in time-varying covariates

- Variables that change over time and captured in your study!
- Difficult to analyze
  - May be issues of missing data that arise
  - Assumptions may be hard to account for
  - Temporality may be questionable (in specific circumstances)
  - Computationally intensive!
- Only appropriate with the right research question
- Are these trajectories affected by other variables that change over time?
  - Do these events alter the trajectory itself?
  - Should they be modeled as an additional outcome?

MSc Thesis: Identifying quality of life trajectories after a prostate cancer diagnosis

- **Megan Farris**,[1,2] Karen Kopciuk,[1,3,4] Kerry Courneya,[5] Elizabeth McGregor,[6] Qinggang Wang,[1] Christine Friedenreich[1,2,3]

1. Department of Cancer Epidemiology and Prevention Research, CancerControl Alberta, Alberta Health Services
2. Department of Community Health Sciences, Cumming School of Medicine, University of Calgary
3. Department of Oncology, Cumming School of Medicine, University of Calgary
4. Department of Mathematics and Statistics, University of Calgary
5. Faculty of Physical Education and Recreation, University of Alberta
6. Alberta Cancer Prevention Legacy Fund, CancerControl Alberta, Alberta Health Services

- MSc Thesis Defense, August 10th, 2016, Calgary, AB
- Work published in: International Journal of Cancer

Burden of Prostate Cancer

- Prostate cancer second most common cancer worldwide in men

- Five-year survival rates between 80-95% in industrialized countries
  - 1990s Prostate Specific Antigen (PSA) screening program

- Early detection and over-detection

- Consequences of over-treatment lead to:
  - Long-term residual side effects
  - Reduced functioning
  - Compromised mental state

- Overall reduced quality of life (QoL) often occurs

Bray F RJ, 2013; Ferlay J S, 2013

Measures of Prevention through Physical Activity

- Many prognostic non-modifiable risk factors for prostate cancer

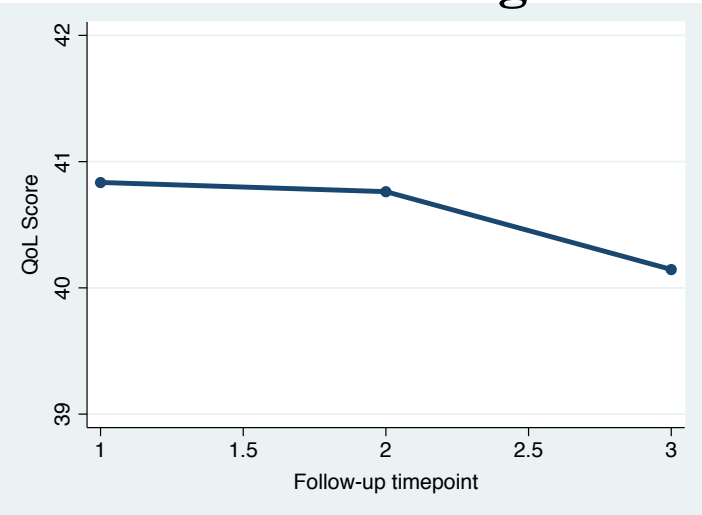
- Physical activity is a modifiable behaviour
  - Promotes overall health status
  - Known to increase QoL in healthy individuals

- Many studies have examined physical activity and how it relates to QoL.
  - Cross-sectional/short-term

Patterns in QoL

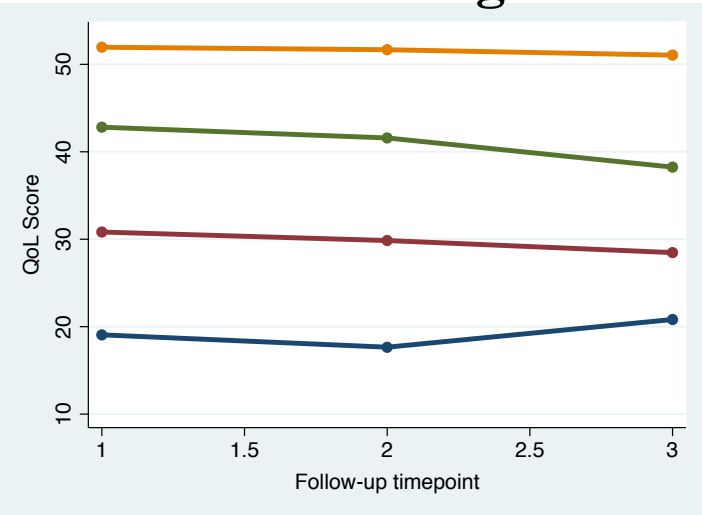
- QoL outcomes are well-documented after diagnosis of prostate cancer

Traditional Longitudinal Modelling



- Assumes one average QoL for population
- Miss subtle differences?

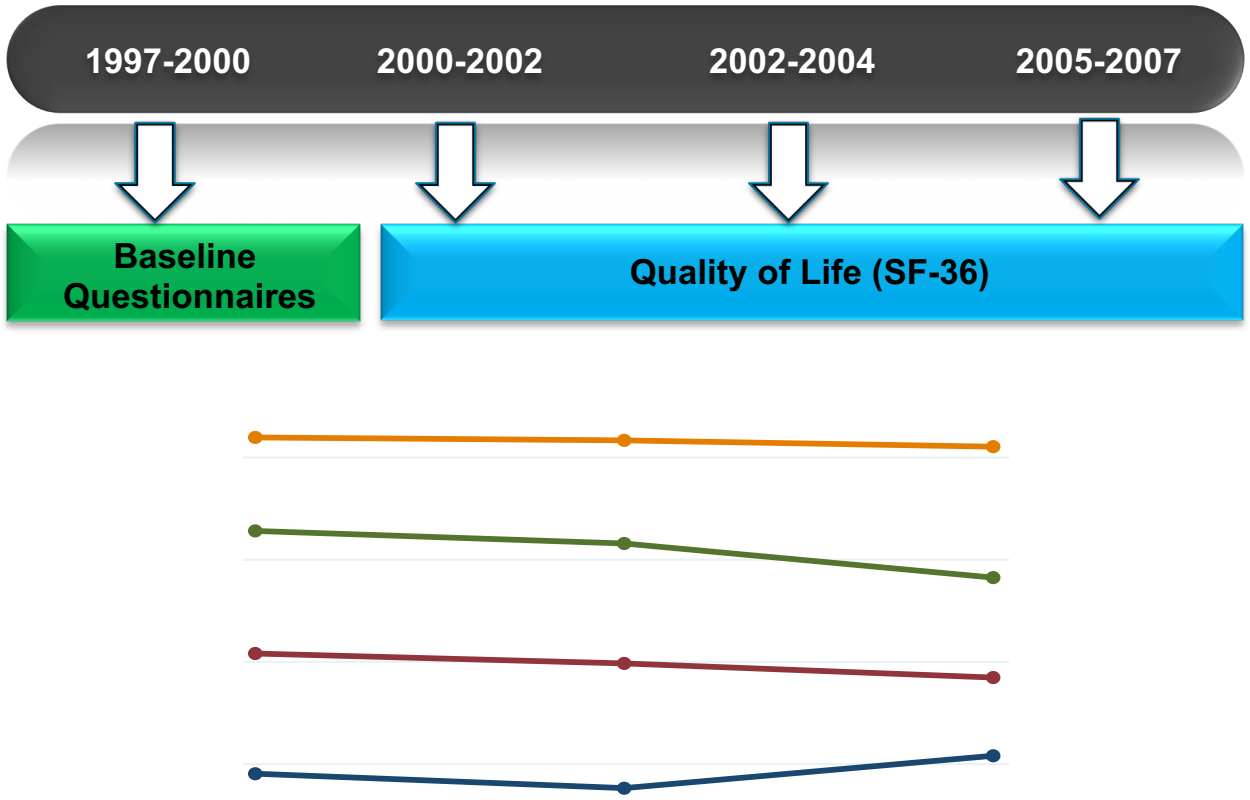
Group-Based Trajectory Modelling



- No distribution assumptions
- Data determines trajectory groups that might be found

Objectives

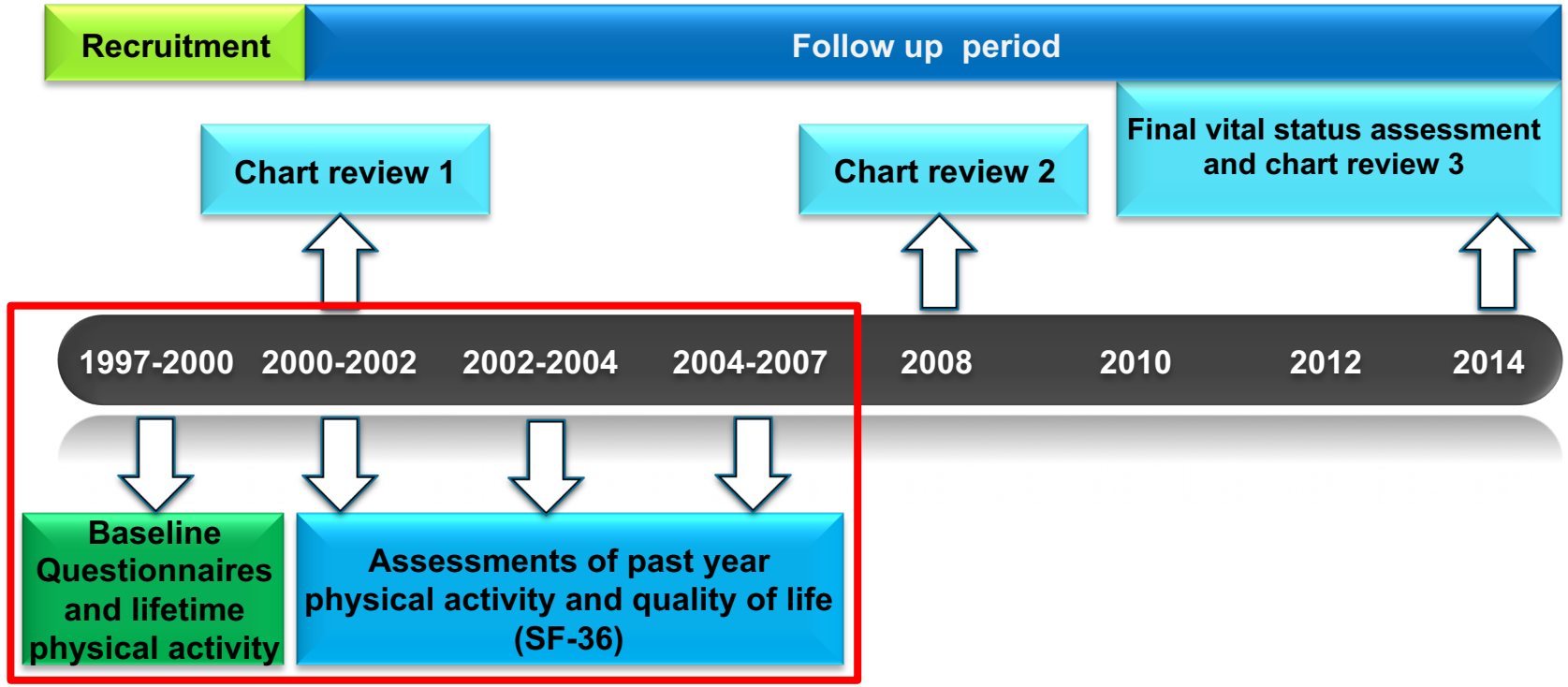
- To examine post-diagnosis QoL trajectory groups in a cohort of prostate cancer survivors during the follow-up period



Study Population/Inclusion Criteria

- Stage T2 or greater, identified through Alberta Cancer Registry
- Under the age of 80 years
- No previous cancer diagnosis

Data Collection Timeline



- Quality of life was collected from the SF-36 self-reported questionnaire
- Physical component summary (PCS) and mental component summary (MCS) scores were derived

Data Collection

- QoL was collected from the SF-36 self-reported questionnaire
  - Collected by self-report questionnaire (all time-points)
  - Validated questionnaire
  - Measures general QoL
  - All eight domain scores of the SF-36 needed to be present
  - Physical component summary (PCS) and mental component summary (MCS) scores were derived

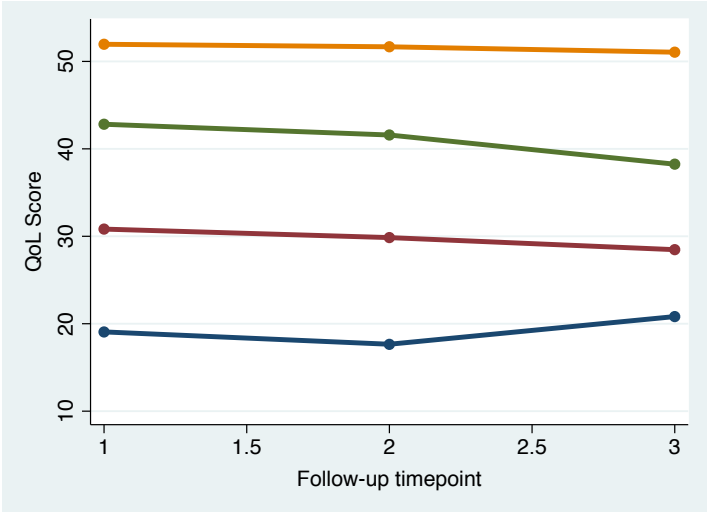
- Physical activity was collected by the lifetime total physical activity questionnaire and past-year total physical activity questionnaires
  - LTPAQ collected by interview
  - PYTPAQ collected by interview (1st follow-up) and questionnaire (2nd and 3rd follow-up)
  - Reliable and validated questionnaire
  - Measures all types, intensities, durations and frequencies

Data Collection

- Interviews from prior case-control study collected:
  - Personal health history
  - Prostate cancer related variables
  - Lifestyle behaviours and study characteristics
  - Anthropometric measurements

Statistical Analysis

- Group-based trajectory modelling are finite mixture models used to approximate unknown distributions of physical and mental QoL trajectories
- Influential dropout was examined
- Fit behaviours and prognostic factors
- Sensitivity analyses were carried out with complete follow-up assessments and time-lagged models for previous QoL score



Traj for STATA®

Secondary Statistical Analysis: Multinomial Logistic Regression

- Influential dropout was examined according to model fit and dropout probabilities between groups

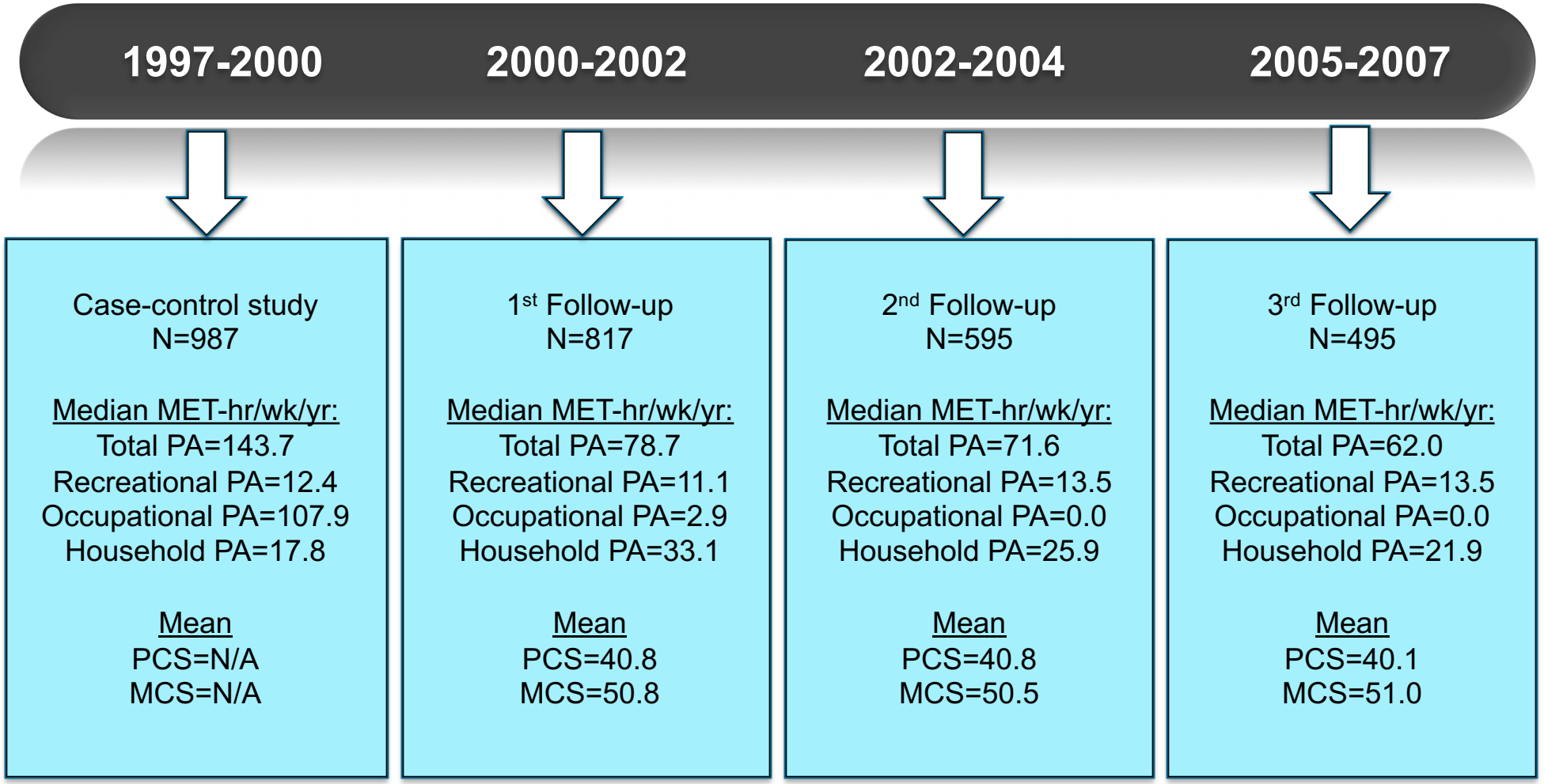
- Fit behaviours and prognostic factors including:
  - Age at diagnosis (years)
  - Aggressive vs. non-aggressive disease (Gleason score ≥ 8, stage > II)
  - Radiation therapy
  - Hormone therapy
  - Prostatectomy
  - Post-diagnosis Charlson co-morbidity score
  - Body mass index (BMI) kg/m^2
  - Smoking status at diagnosis

- Sensitivity analyses were carried out with complete follow-up assessments and time-lagged models for previous QoL score

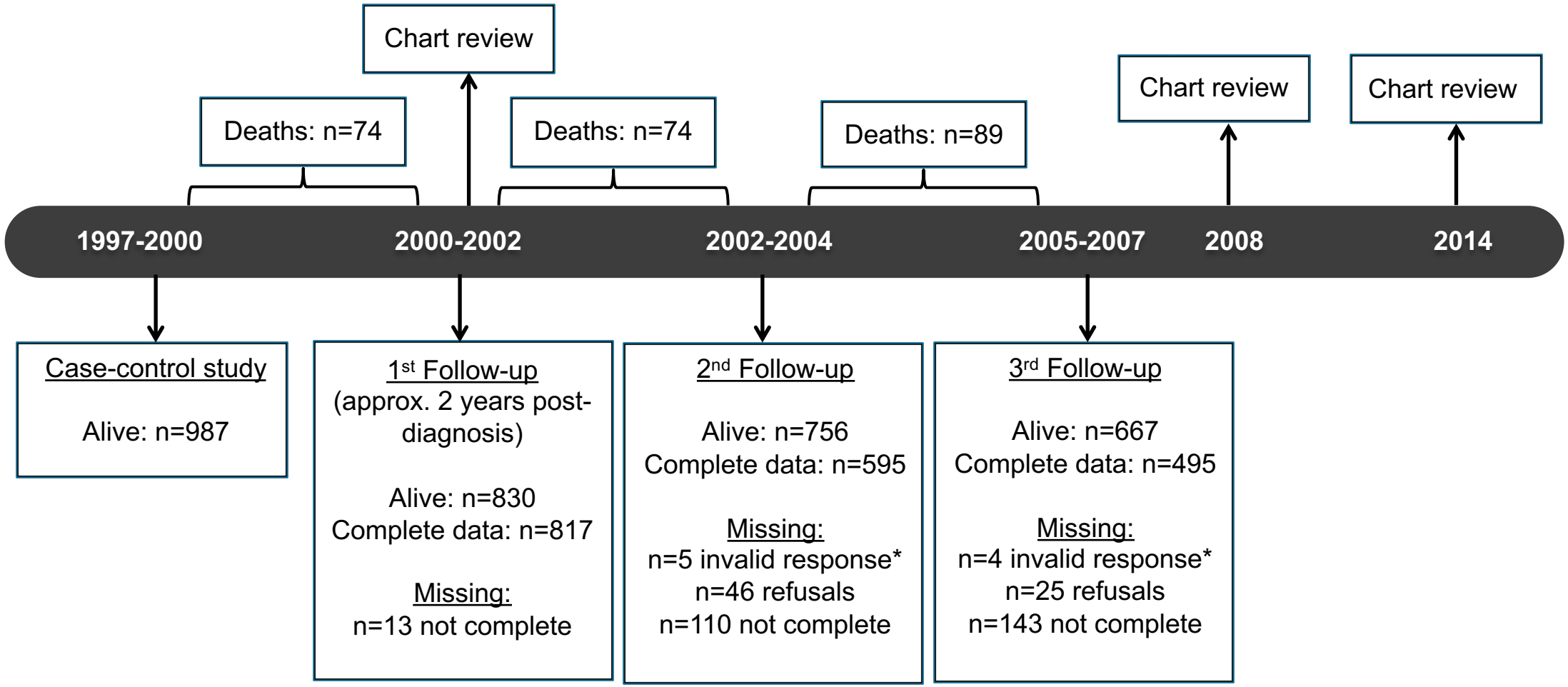
Characteristics of Prostate Cancer Survivors in Alberta

Characteristics	Total sample
	N (%)
Stage of cancer	
II (T1/T2, N0, M0)	630 (77.1%)
III (T3, N0, M0)	57 (7.0%)
III/IV (T3, NX, MX)	75 (9.2%)
IV	55 (6.7%)
Primary Treatment	
Prostatectomy	240 (29.4%)
Hormone therapy	517 (63.3%)
Radiation therapy	359 (43.9%)
Relationship status	
Married/ common law	689 (84.3%)
Other	128 (15.7%)
	Mean (SD)
Age at diagnosis (years)	67.3 (7.4)
Body mass index (kg/m^2)	28.0 (3.8)

Characteristics of Prostate Cancer Survivors in Alberta



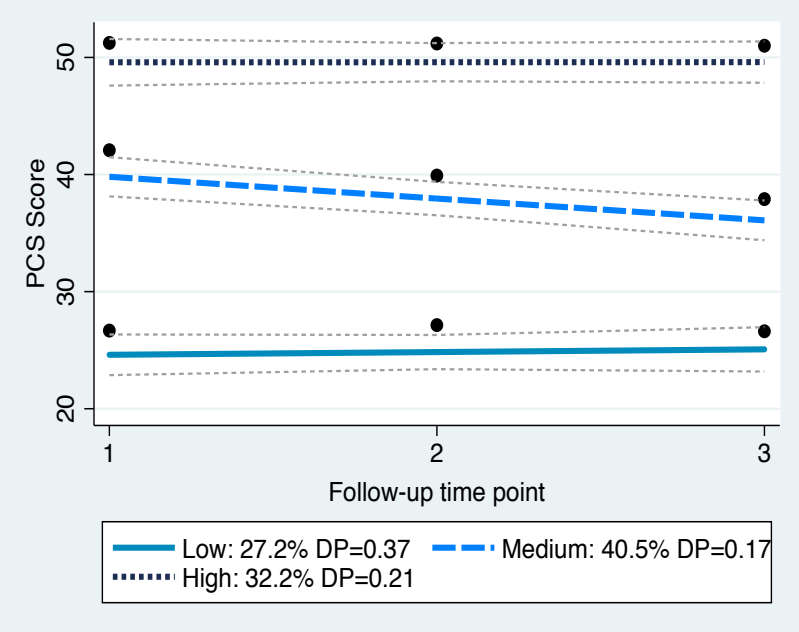
Reasons for loss to follow-up data collection timeline



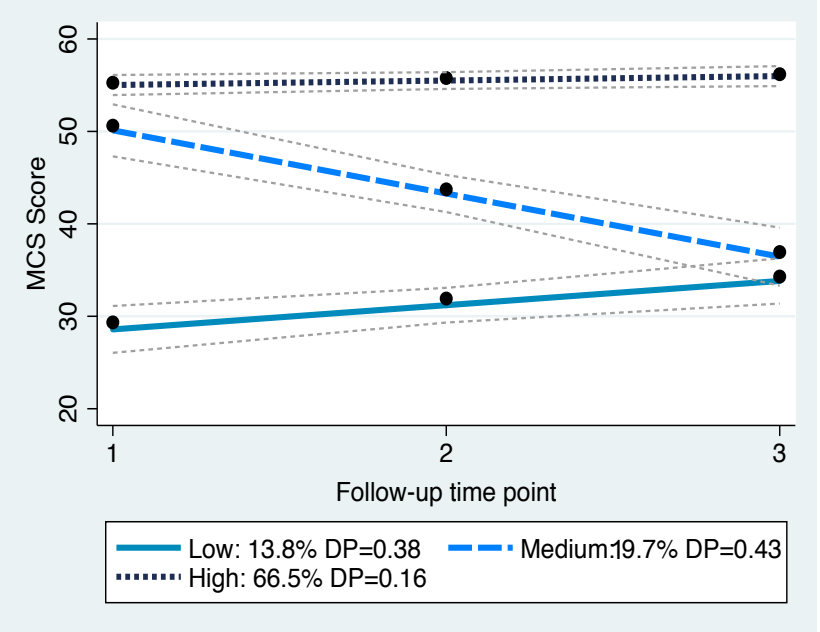
*One or more SF-36 domain scores were missing

Group-Based Trajectory Model Results

Physical QoL



Mental QoL



Characterizing Trajectories

Prognostic/behavioural factors	Medium versus High QoL		Low versus High QoL	
	RRR [a]	95% CI	RRR [a]	95% CI
Physical QoL trajectories				
Age at diagnosis (years)	1.04	0.99-1.08	1.11	1.06-1.16
Aggressiveness of disease [b]	1.41	0.72-2.79	1.50	0.76-2.95
Radiation therapy [b]	1.10	0.57-2.12	0.64	0.35-1.17
Hormone therapy [b]	1.76	1.02-3.02	1.99	1.14-3.48
Prostatectomy [b]	1.25	0.58-2.68	2.29	1.07-4.89
Charlson co-morbidity score	1.59	1.29-1.95	2.04	1.66-2.51
BMI (kg/m^2)	1.09	1.02-1.17	1.18	1.09-1.27
Smoking status [b]	1.71	0.75-3.90	2.77	1.23-6.26

[a] All models were adjusted for time-varying physical activity, dropout probabilities and all other factors in tables
[b] Dichotomous variables

Characterizing Trajectories

Prognostic/behavioural factors	Medium versus High QoL		Low versus High QoL	
	RRR [a]	95% CI	RRR [a]	95% CI
Mental QoL trajectories				
Age at diagnosis (years)	1.06	1.01-1.11	1.00	0.97-1.04
Aggressiveness of disease [b]	0.89	0.44-1.81	0.40	0.19-0.86
Radiation therapy [b]	0.57	0.30-1.10	0.50	0.28-0.89
Hormone therapy [b]	1.61	0.82-3.16	1.01	0.61-1.67
Prostatectomy [b]	3.32	1.23-8.94	1.73	0.89-3.36
Charlson co-morbidity score	1.32	1.12-1.56	1.18	1.03-1.36
BMI (kg/m^2)	1.04	0.96-1.12	1.04	0.97-1.10
Smoking status [b]	1.79	0.75-4.24	2.35	1.27-4.36

[a] All models were adjusted for time-varying physical activity, dropout probabilities and all other factors in tables
[b] Dichotomous variables

Mean QoL Coefficients & Standard Errors

Trajectory groups	Baseline intercept	Slope	Physical activity slope
Physical quality of life trajectories			
High-maintaining	49.64 (1.38) [a]	-0.03 (0.49)	0.02 (0.01) [b]
Medium-declining	41.68 (1.16) [a]	-1.85 (0.46) [a]	0.02 (0.01) [a]
Low-maintaining	24.37 (1.28) [a]	0.19 (0.55)	0.03 (0.01) [b]
Mental quality of life trajectories			
High-increasing	54.59 (0.76) [a]	0.45 (0.30)	0.01 (0.01)
Medium-declining	57.32 (2.37) [a]	-6.98 (1.13) [a]	0.01 (0.01)
Low-increasing	25.99 (1.97) [a]	2.62 (0.84) [b]	0.01 (0.01)

[a] p-value < 0.001.
[b] p-value < 0.01.

Things to keep in mind

- This is a newer technique that may be very powerful for estimating causal inference in Epidemiological studies
- Still an evolving analysis method
- Follow his recommended reporting framework (applicable to all analysis types)
  - Rationale and transparency are key!!
- Be mindful of limitations this type of analysis might have (due to assumptions etc.)

Conclusion

- First study to detect three distinct groups of physical and mental QoL up to 10 years post-prostate cancer diagnosis
- Confirmation of these findings is warranted in this population
- Characteristics of QoL trajectories will lead to a better understanding of differences between groups and how health professionals and researchers can use this information

Acknowledgements

Supervisory committee:
- Dr. Christine Friedenreich
- Dr. Kerry Courneya
- Dr. Karen Kopciuk
- Dr. Elizabeth McGregor

Study staff: Aleata Rhyorchuck, Sana Fakih, Jodi Parrotta, Linda Davison, Pearle Cooke, Nicole Slot, Carol-Anne Zawalykut, Catherine Munro, Carla Quesnel, Yvonne LeBlanc, Sarah MacLaughlin and Eileen Shaw

CIHR IRSC
Canadian Institutes of Health Research | Instituts de recherche en santé du Canada

UNIVERSITY OF ALBERTA

Alberta Health Services

Institut national du cancer du Canada | National Cancer Institute of Canada

UNIVERSITY OF CALGARY

Thank you!

Any questions?

Megan.farris11@gmail.com