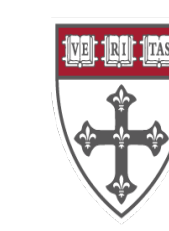


Methods and Software for Penetrance Estimation in Complex Family-based Cancer Penetrance Studies



Dana-Farber
Cancer Institute



HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Nicolas Kubista^{1,2}, Stephen B. Gruber³, Gregory E. Idos³, Danielle Braun^{1,2}, Giovanni Parmigiani^{1,2}.

¹Dana-Farber Cancer Institute, ²Harvard University T.H. Chan School of Public Health, ³City of Hope National Medical Center

Background

What is Penetrance?

Penetrance is the proportion of individuals carrying a genetic variant who develop associated symptoms by a given age. For cancer genetics, it represents the probability that a mutation carrier will develop cancer at a specific age.

Why It Matters

- Critical for accurate cancer risk assessment for cancer syndromes with a genetic heritability component
- Guides clinical decision-making and prevention strategies (e.g. enhanced cancer screening, preventive surgical intervention)
- Essential for genetic counseling and patient care in clinics

Current Challenges

- Limited software exists for age-specific estimation.
- Currently no Bayesian estimation frameworks exist that allow the incorporation of learnings from prior studies.
- Complex family structures difficult to analyze.
- Significant missingness in many studies.

Methods

1 Genetic Model: Mendelian inheritance

- Input data are proband-centered pedigrees, usually with limited genetic testing information where only the proband has been genotyped but relatives' carrier status is unknown
- Format is consistent with *progeny*

2 Bayesian Estimation Framework

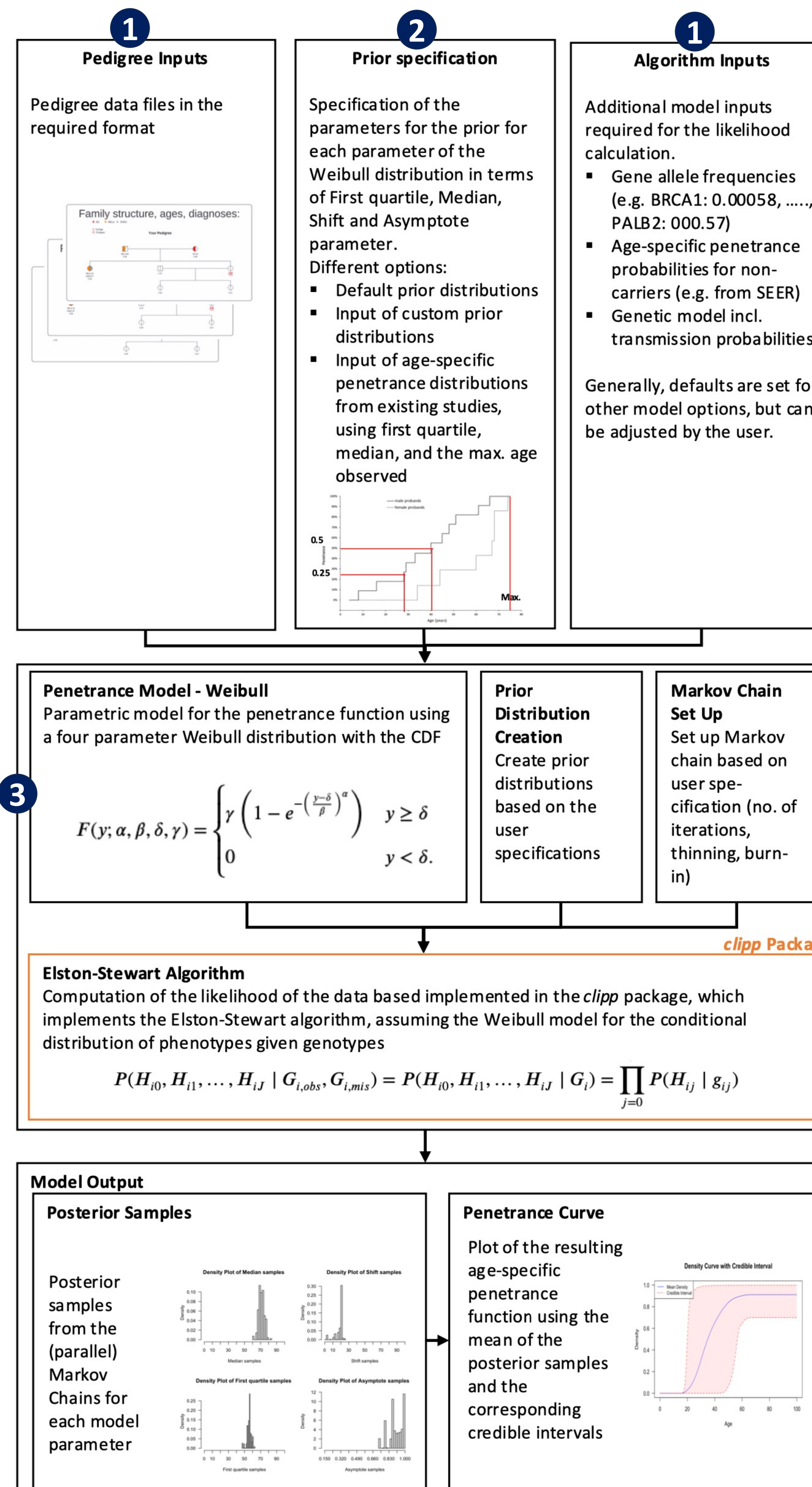
- We allow the user to specify priors on the parameters of the four-parameters
- Estimation of the model parameters using an Adaptive Metropolis-Hastings MCMC
- Uncertainty quantification via posterior sampling

3 Weibull Parameterization

- We model the time-to-event data (where the event is the cancer diagnosis) using a parametric time-to-event model with a modified four-parameter Weibull distribution to model the penetrance function (see graphic)

Methods

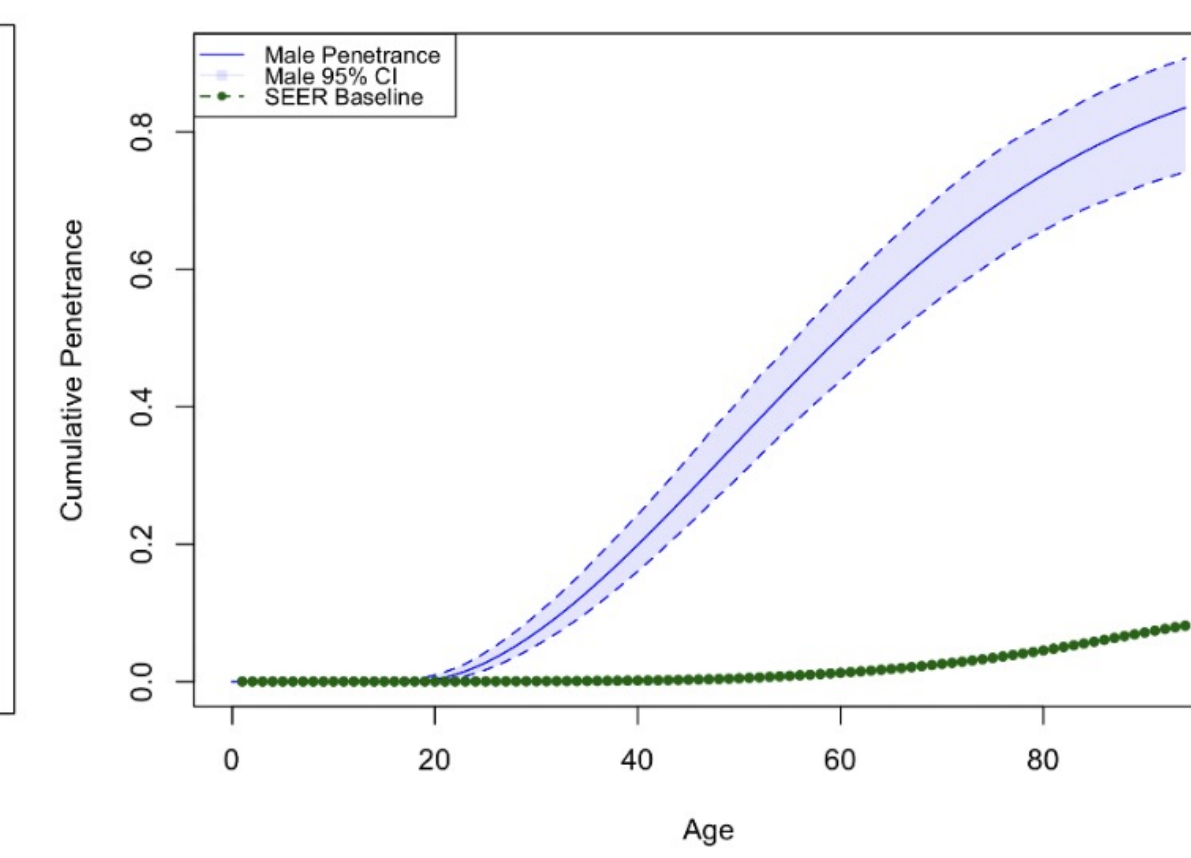
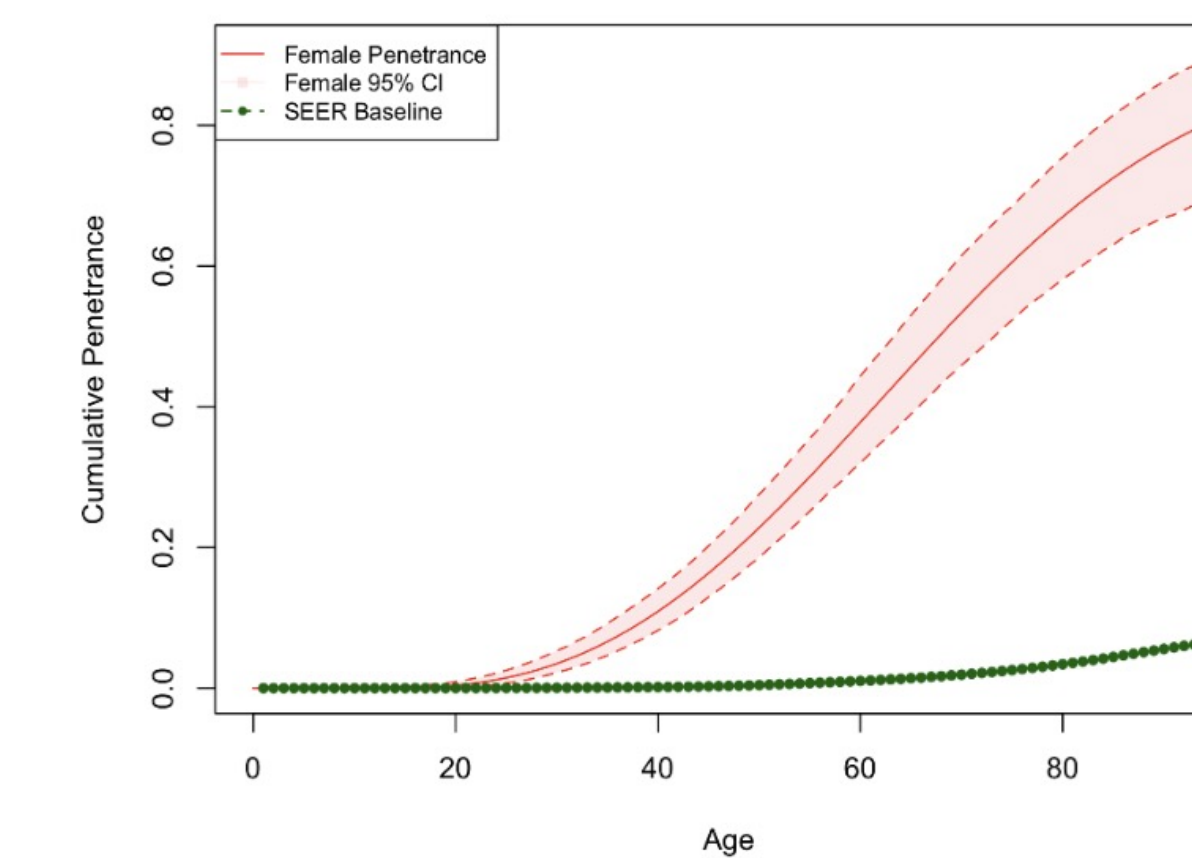
The *penetrance* package workflow



Methods

Age-specific Penetrance Estimation for Colorectal Cancer and the MLH1 gene

- Application of *penetrance* to a recent dataset from the Clinical Cancer Genomics Community Research Network (CCGRN) with 130 families (4,551 individuals), where the proband was a carrier of the MLH1 PGV, and 486 colorectal cancer diagnoses
- *Penetrance* directly produces age-specific penetrance curves as an output



- Estimates show significantly increased life-time risk for colorectal cancer for both females and males who are carriers of the MLH1 PGV
 - Males: 65% cumulative risk by age 70 [95% CI: 57-74%]
 - Females: 56% cumulative risk by age 70 [95% CI: 49-64%]

Conclusions

- *Penetrance* provides an important tool for cancer risk prediction and offers significant advantages over the currently available software for genetic risk estimation:
 - Robust Bayesian framework for uncertainty quantification
 - User-friendly implementation in an R package
 - Automatic treatment of missing genotype data
 - Option for age-imputation
- The software package has already received positive feedback from clinicians and is currently being applied in the largest study on Li-Fraumeni Syndrome and TP53

Acknowledgements: the authors would like to acknowledge support from the National Cancer Institute at the National Institutes of Health grants R01CA242218.