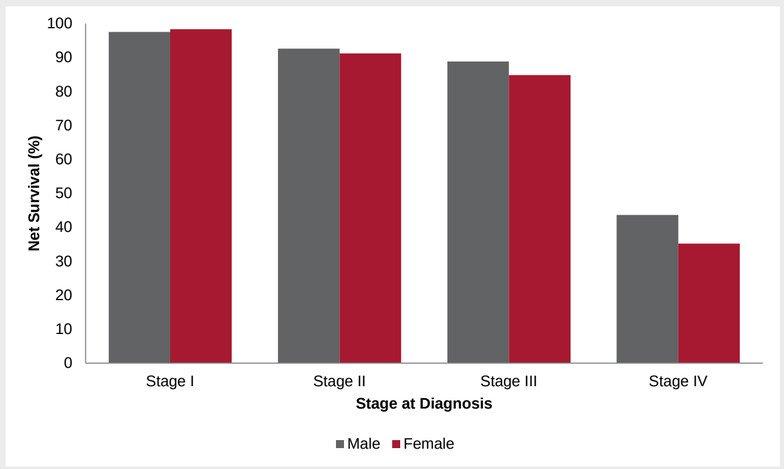
Genetic and non-genetic risk modelling in colorectal cancer

Philip J. Law, Richard S. Houlston

Introduction

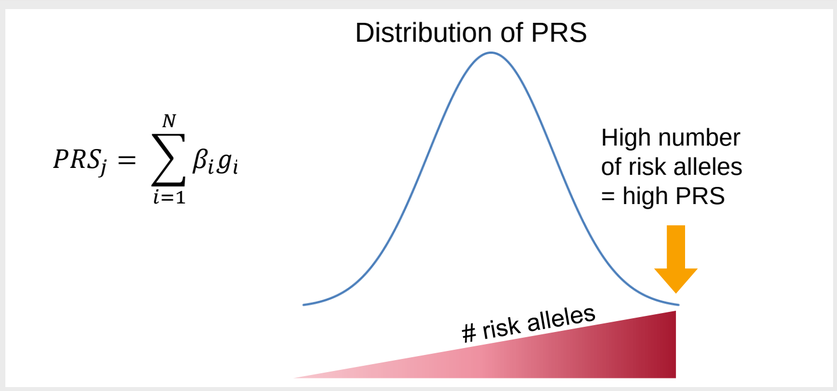
- In the Western countries, colorectal cancer (CRC) affects over half a million individuals each year.
  - Age standardised incidence rate ~70/100,000 cases in UK in 2015 (CRUK)
  - 5-year survival remains only ~55%.
  - Strong correlation between stage and survival
  - Thus, early detection is vitally important



CRC 1-Year Age Standardised Net Survival by Stage, Adults (Ages 15-99 Years), England 2014, CRUK

- Multiple risk factors:
  - Lifestyle factors
    - Smoking
    - BMI
    - Processed meat
    - Alcohol consumption
    - Hormone therapy
    - Physical activity
    - Fruit and vegetable consumption
  - Genetic factors
    - Family history
    - High penetrance mutations, e.g. *APC, MLH1*, *MSH2*
    - Common, low penetrance variation
- Current screening is based on:
  - Faecal occult blood test (FOBT)
  - Faecal immunochemical test (FIT)
  - Colonoscopy
- Screening generally age based
  - NHS guidelines recommend screening for individuals older than 55, and regular screening for individuals 60-74
- Effectiveness of screening and prevention programs may be optimised when directed toward those individuals at highest risk
  - Stratified screening for those with higher genetic risk

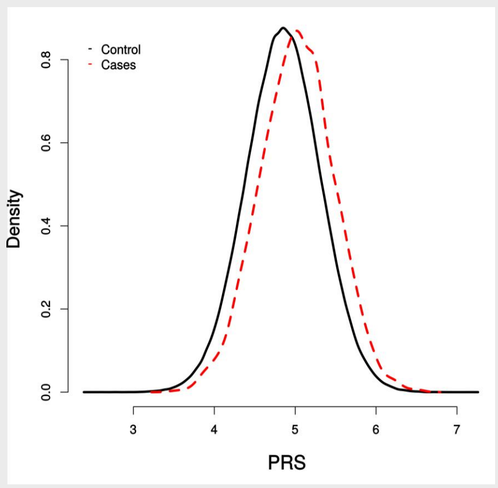
Polygenic Risk Score



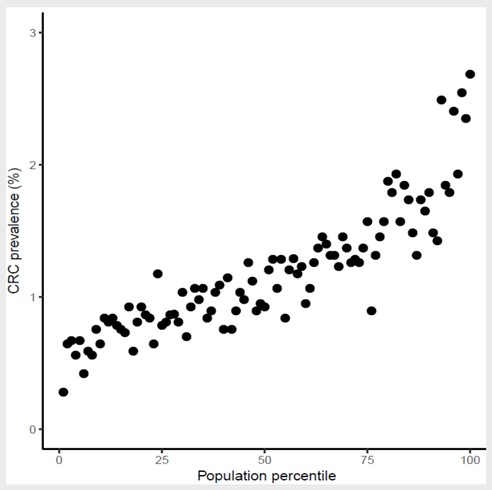
- Polygenic risk scores (PRS) provide a simple composite metric with which to interpret a genetic profile
  - Calculated for each individual *j* as a sum of allele dose for single nucleotide polymorphisms (SNP) *i*, weighted by the per-allele log-odds ratio (effect size) for that SNP (β_i)
  - *N* is the number of SNPs used. Genome-wide association studies (GWAS) have identified over 50 risk variants strongly associated with CRC risk
  - Hypothesis: individuals with more risk alleles will have a higher PRS
- We have previously investigated a PRS based screening approach for CRC, as compared to standard age-based screening
  - Potential to result in 26% fewer individuals requiring screening, at the cost of 6% fewer screen-detected cases.

Results

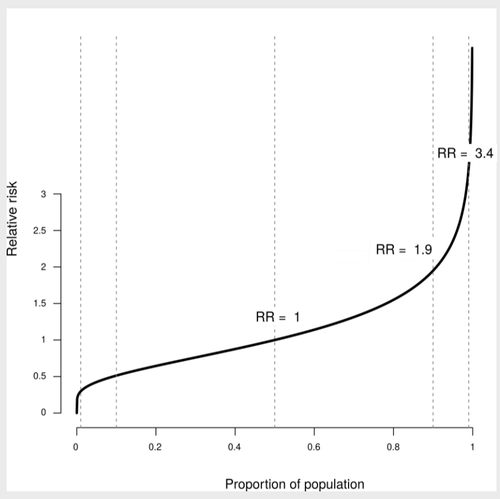
Using the UK Biobank, we identified 4,225 individuals with primary diagnosis of CRC, and 353,225 cancer-free controls.



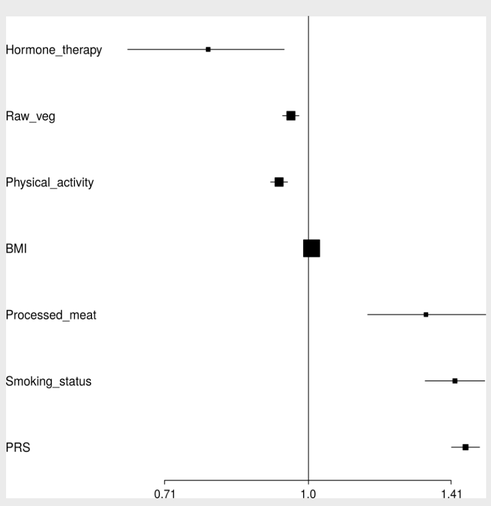
Distribution of PRS in cases and controls, showing a right-shift in the scores in the individuals with CRC (red dotted line) as compared to cancer-free individuals (solid black line).



The prevalence of CRC increases when stratifying the dataset into PRS percentiles (~3,575 individuals in each percentile).



Population distribution of PRS ordered by relative risk (compared with population median risk). Vertical lines (left to right) correspond to 1%, 10%, 50%, 90%, and 99% centile, respectively. Individuals in the top 1% of PRS have a 3.4 increased risk compared to the population median.



With the UK Biobank, we have access to a multitude of lifestyle factors. Including the known risk factors into a logistic regression confirms these factors. Figure shows a forest plot of these risk factors. Mid-line represents OR=1.0.

PRS allows the identification of individuals with a higher risk of CRC, highlighting those who may benefit from earlier or more regular screening.

Future work

The UK Biobank contains data on a multitude of lifestyle factors. Using integrative techniques, we aim to build predictive models that combine the polygenic risk scores together with lifestyle data to identify modifiable risk factors.

We will implement this within a machine learning strategy, through the investigation of a number of approaches. Feature selection will allow the identification of significant risk factors.

Validation will be pursued through established collaborative links with screening agencies including National Bowel Cancer Hubs and other cohort studies, including the Breast Cancer Now Generations Study.

These models have the potential to optimise population screening for CRC, and define those individuals most likely to benefit from chemopreventative agents.

References

Frampton *et al*. Implications of polygenic risk for personalised colorectal cancer screening, Ann Oncol, 2016

Johnson *et al.* Meta-analyses of Colorectal Cancer Risk Factors. Cancer Causes Control, 2013

Botteri *et al.* Smoking and Colorectal Cancer: A Meta-analysis. The Journal of the American Medical Association, 2008

Contact

Molecular & Population Genetics
Division of Genetics & Epidemiology

Philip.Law@icr.ac.uk

Acknowledgments

Amit Sud
UK Biobank

@HoulstonLab_ICR