Assessing the susceptibility of celiac disease by polygenic risk scores: analysis of a population-based cohort, the HUNT study.

Mohammad Sayeef Alam[1,2], Rebecka Hjort[1,2], Kristian Hveem[1], Knut E. A. Lundin[3,4], Iris H. Jonkers[5], Ludvig M. Sollid[4,6], Eivind Ness-Jensen[1,2,7,8]

1 HUNT Center for Molecular and Clinical Epidemiology, Norwegian University of Science and Technology, Trondheim, Norway.
2 HUNT Research Centre, Norwegian University of Science and Technology, Levanger, Norway.
3 Department of Gastroenterology, Oslo University Hospital-Rikshospitalet, Oslo, Norway.
4 Norwegian Coeliac Disease Research Centre, Institute of Clinical Medicine, University of Oslo, Oslo, Norway.
5 Department of Genetics, University of Groningen, Groningen, The Netherlands.
6 Department of Immunology, Oslo University Hospital-Rikshospitalet, Oslo, Norway.
7 Department of Medicine, Levanger Hospital, Nord-Trøndelag Hospital Trust, Levanger, Norway.
8 Department of Molecular Medicine and Surgery, Karolinska Institutet & Karolinska University Hospital, Stockholm, Sweden.

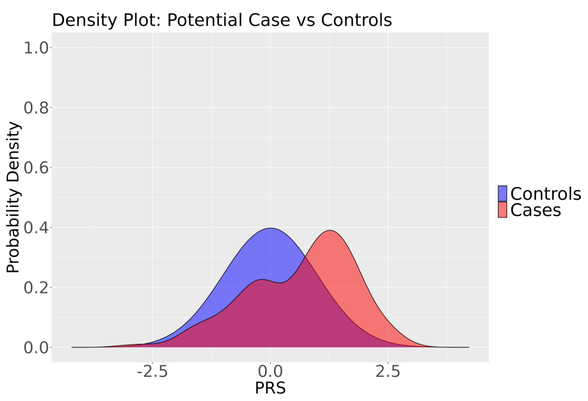
This is what our issue is:

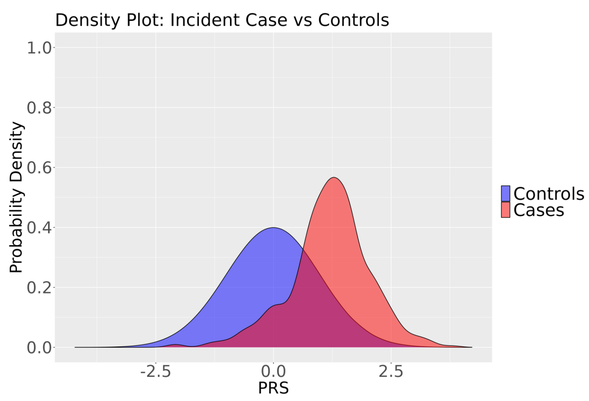
1. **Celiac disease (CeD) is a complex disease triggered by gluten consumption in genetically susceptible individuals.**
2. **The disease can onset at any age with a wide range of symptoms, often leading to undiagnosed cases due to symptom overlap with other conditions or lack of awareness.**
3. **Current diagnostic methods involve repeated serological examinations and invasive biopsies, with a strict gluten-free diet recommended upon diagnosis.**
4. **Genetic screening for HLA DQ2/DQ8 haplotype is recommended for high-risk patients, but these tests have a low positive predictive value.**
5. **The use of a polygenic risk score (PRS) is encouraged to assign a probability of disease development based on an individual's genetic makeup, aiding in early intervention and potentially improving patient outcomes.**

To investigate it, we did this:

1. **The study used data from the HUNT study, conducted in Nord-Trøndelag County, Norway between 2017-2019, with a focus on adults over 19 years old, yielding a 54% response rate and 56,939 participants, over 50% of whom were women.**
2. **A serological was used to test for both TG2 IgA and IgG antibodies. Confirmed cases of CeD were identified based on positive serology and biopsy results, with incident cases being those identified during the study and prevalent cases being those with a previous diagnosis.**
3. **Genotyping was performed using Illumina HumanCoreExome arrays, and imputation was carried out using the Haplotype Reference Consortium (HRC) reference panel and the Positional Burrows Wheeler Transform method.**
4. **Stratification of HLA risk groups was based on imputed HLA genotypes*, with individuals classified into low-risk, high-risk, and intermediate risk groups based on their HLA-DQ2/DQ8 status. Similarly, the PRS was stratified into three groups. (***data provided by NIDDK CR, a program of the National Institute of Diabetes and Digestive and Kidney Diseases.**)**
5. **The study reproduced a previously developed PRS with the highest accuracy, incorporating a Bayesian regression framework for marker-specific adaptive shrinkage.**
6. **The overall discrimination of the genetic model was evaluated by calculating the area under the ROC curve for HLA only and combining HLA and the PRS, as well as the net reclassification improvement and the integrated discrimination improvement.**

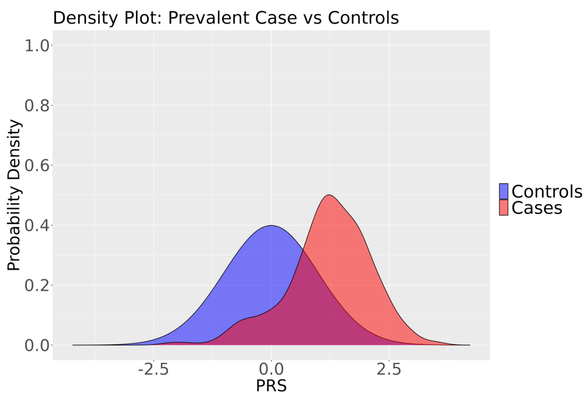
We found out that:

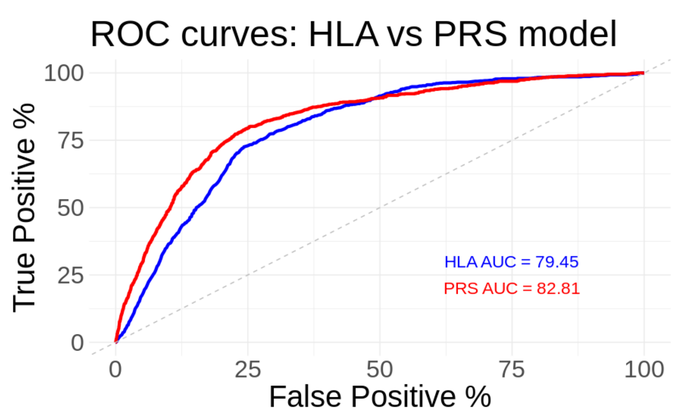




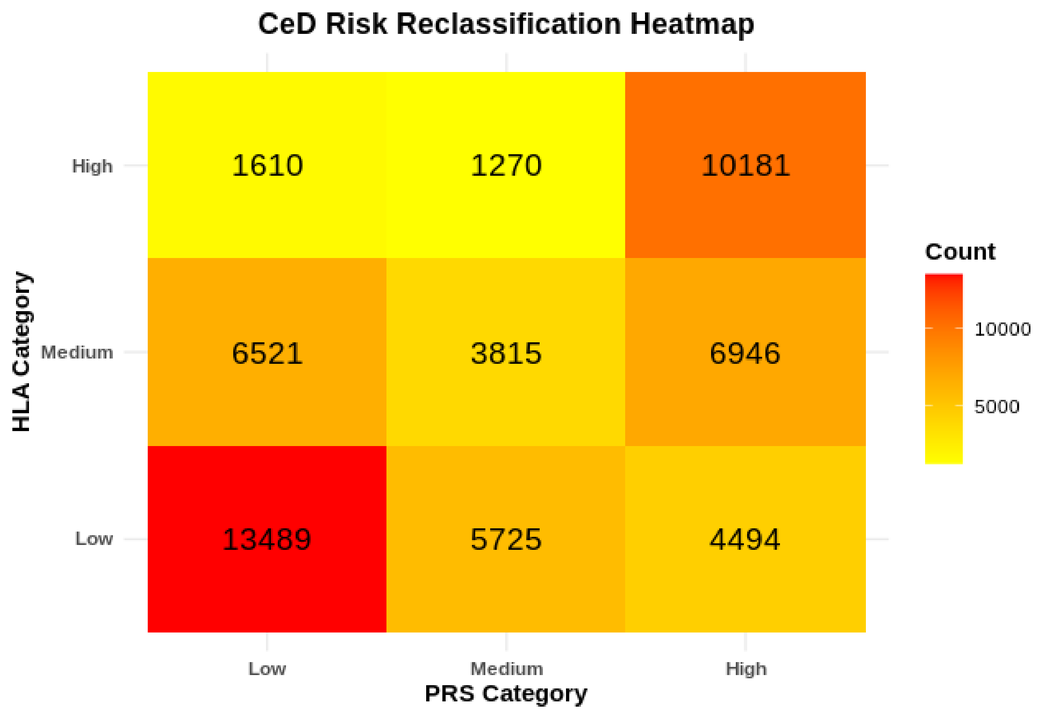
The mean PRS score for prevalent cases, incident cases as well as potential cases are higher than the controls.

Results (contd):





The ROC curve indicating the classification of confirmed cases accurately shows that the PRS performs slightly better with an AUC of 82.81% compared to 79.45%.



From the above reclassification heatmap we can draw our focus the 4494 who are in the top quartile of PRS but according to their HLA are in a low-risk category. Similarly, individuals with medium or low HLA risk and PRS risk could altogether be chosen not to be screened saving cost and time.

Our take home message is that:

1. **The PRS assigns a probabilistic value of developing CeD, with individuals in the top quintile at a higher risk. However, it doesn't predict the final outcome and its performance in differentiating potential cases from controls is poor.**
2. **Removing the HLA variants from the PRS decreases its discrimination power, indicating that the major discrimination power between potential and confirmed cases lies with the HLA variants included in the PRS.**
3. **The PRS can reclassify a significant percentage of individuals from the lower HLA risk group to a high PRS risk group, identifying populations at risk beyond HLA screening.**
4. **The PRS could potentially be incorporated during newborn screening, saving the healthcare sector from unnecessary screening and testing, and sparing patients from discomfort and time if they are less likely to develop the disease.**
5. **Despite the limitations of the model, including non-HLA risk factors can reclassify a significant portion of the population into more accurate risk categories, which might help to make a better selection of those who need closer follow-up and repetitive antibody testing. This could lead to improved diagnosis of CeD and early intervention, alleviating the socio-economic burden on patients, their families, and society.**