

Yoyo Benchetrit & Anika Arora

March 3, 2024

Predicting Low Lipid Levels for Abetalipoproteinemia (ABL)

The ABL+ Foundation's mission is to provide guidance on needed scientific research, diagnosis, and management of abetalipoproteinemia (ABL) and related hyperlipidemias. Patients with ABL do not produce the protein (apolipoprotein-B) that is necessary to make lipoproteins. It is a deficiency in the MTP gene that revokes assembly of apo-B in the liver and small intestine, resulting in this malabsorption of fat and fat-soluble vitamins.

The underdiagnosis of ABL partly stems from the fact that low levels of lipids (LDL, HDL, cholesterol, and triglyceride) as well as low apolipoprotein-B, are opportunistically tested for. Consequently, potential patients may be simply unaware of their risk of having the disease and never get the genetic testing for MTP deficiency. Generally, children get their lipids profiled by the age of 10. However, there is a need for intervention in ABL earlier on. Thus, our goal in this project was to first provide compelling reason to have a lipid screening program available at birth to have a diagnosis and clinical intervention earlier on in childhood. These types of lipid screens have been denied as the diagnostic criteria is considered too rare. However, we need these screens early on in childhood to aid in underdiagnosis of ABL. In the meantime, if we don't have such screens, we need to figure out other ways to predict these low lipid and apo-B levels – which is what we are tackling in our predictive computational project.

Using data from the National Health and Nutrition Survey (NHANES) (2013-2016), we calculated 2/8612 individuals with LDL levels that are below the threshold to meet diagnostic entry (that is, LDL levels < 15mg/dL). Therefore, we can approximate that $(2/8612)^2 \times 33 \text{ million} = 18$ Americans who have low LDL levels that could correspond to having ABL. Considering that the American population makes up only 4.3% of the world population, these results outline the potential for hundreds of undiagnosed ABL patients. We employed a similar calculation with apo-B levels, in which we calculated that 3/5867 individuals in our data set had low apo-B levels (< 20mg/dL), corresponding to approximately 86 American individuals who may be underdiagnosed. These calculations are just rough approximations to provide reason for diagnosis earlier on, as currently there are only 150 individuals officially diagnosed with ABL.

After outlining this rough sketch of the underdiagnosis of ABL, we used this same NHANES dataset to create a deep-learning model that can predict low lipid (specifically LDL) and apo-B levels from more commonly tested physical measurements and biomarkers. Through aggregating data, we are looking to find relationships between common markers such as blood pressure, height, weight, blood iron levels, etc. that may provide insight into whether someone is at risk of having low LDL (low-density lipoproteins) and/or apoB levels without them having to be tested for these levels directly. We have trained the model to

learn to predict these extreme levels of LDL from such general biomarkers as well as fatty liver disease (which correlates with age), heart disease, anemia, jaundice, and other more commonly known diseases.

Our model is trained on around 3000 samples from the NHANES dataset. Once users submit this form, a prediction of their LDL and apo-B levels is outputted, alerting patients if they are at risk of needing further testing. If patients are at risk, they are provided resources through the web-app and connected with lipidologists near them. By simply entering their location under our resources page, a map with nearby lipidology clinics is offered to encourage patients to see specialists that can confirm low lipid profiles – and hence connect them with genetic testing to confirm diagnosis.

Additionally, to promote the ABL Foundation's goal in spreading awareness of this rare disease, we have added an outreach function to our model that allows for a collection of health information and blood samples to be shared by volunteers. Patients can fill out a questionnaire and user form to collect information such as reduced eye function, muscle weakness, BMI, gastrointestinal issues, deficient immunity, anemia, and more to gather a comprehensive overview of any potential links to low LDL and apo-B levels. In such a way, our model will continue gathering data to make stronger predictions for potential patients and decipher more definitive links between the current diagnostic criteria for ABL, and more common physical and serological biomarkers that may be overlooked. The hope is to collect enough data to eventually have it run through our predictive model with these added features. Our model will also be useful in differentiating between ABL and other related disorders – such as chylomicron retentive disease and familial hypobetalipoproteinemia – as although these related diseases present indistinguishable clinical features, patients differ in their lipid level profiles.

We have developed this predictive model on a cohort of adults, primarily because we had access to this data. Even with a small dataset, the model's accuracy in predicting LDL levels is 75% and apo-B is 80%. Thus, the model has been useful in beginning to understand which variables may be predictive of low LDL and apo-B, but we need to continue analyzing data and bring this information to the pediatric population if we want to employ early intervention. That is, the older population that have had screening have well-documented lipid profiles that we can use as a cohort to characterize relating variables when trying to generate universal biomarkers/physical measurements that correlate to low LDL and apo-B levels in children. The next steps will involve validating and generalizing the most predictive features of our model to the general population of children who experience a failure to thrive with ABL.

In the future, we hope to have a module component specifically for ABL patients to share data to aid in predicting the more extreme low levels of LDL and apo-B that manifest in individuals with this disease. As the ABL+ Foundation strongly emphasizes bringing awareness and providing access to patient communities, we have focused on developing a model that has this dual function in predicting diagnosis and encouraging outsourcing and community involvement to aid in the underdiagnosis of ABL. By potentially partnering with

further resources such as Mount Sinai's genome sequencing of large number of adult patients, and their access to very sensitive medical records, we will be able to find more specific populations/cohorts to validate and generalize our predictive tool to be more applicable to earlier diagnosis and intervention.