SUPPLEMENTAL MATERIAL

Raitakari *et al*. **Lipoprotein (a) in youth and prediction of major cardiovascular outcomes in adulthood.**

1. Extended Methods

2. Extended Results

1. Extended Methods

1.1 Description of the Cardiovascular Risk in Young Finns Study (YFS)

The YFS is a population-based prospective follow-up study on cardiovascular risk factors in Finland. It has been carried out in all five Finnish university cities with medical schools and their rural surroundings. The first cross-sectional study was conducted in 1980. Altogether 4,320 children and adolescents aged 3, 6, 9, 12, 15 and 18 years were randomly recruited from the population register of these areas to produce a representative subsample of Finnish children. Of these individuals 3,596 (83%) participated in year 1980. Since then, regular follow-up visits have been performed in 1980, 1983, 1986, 2001, 2007, 2011 and 2018/2020. Follow-up using registry data (diagnoses, medications and mortality) have been extended by 2018. The YFS has been approved by the Joint Commission on Ethics of the Turku University and the Turku University Central Hospital and has been conducted according to the guidelines of the Declaration of Helsinki, and informed consent has been obtained from all participants or their parents.

1.2 Physical examination in YFS

Height was measured using a wall-mounted Seca stadiometer with 0.5 cm accuracy, and weight was measured with a digital Seca scale to the nearest 0.1 kg. Body mass index (BMI) was calculated from the formula: BMI =weight (kg)/height (m). Blood pressure (BP) was measured from the right brachial

artery with a standard mercury sphygmomanometer in 1980 and 1983, and with a random-zero sphygmomanometer (Hawksley & Sons Ltd, Lancin, UK) in 1986, 2001, 2007 and 2011. BP was measured in the sitting position after a 5-minute rest. The Korotkoff first phase was used as the sign of SBP, and DBP was determined from the fifth phase. Readings were performed to the nearest integer of millimeters of mercury 3 times on each participant and the average of 3 measurements was used in the analysis. Participants aged 12 years or older were asked to fill out a questionnaire on smoking. Youth smoking status was dichotomized into daily smokers and non-smokers, defined as current daily smoking (yes/no) at baseline or at any of the follow-up studies when the participants were age 12 to 24 years. Parental education was queried in questionnaires.

In study years 1980, 1983, 1986, 1989, and 1992, venous blood samples were drawn after a 12-hour fast from the right antecubital vein with the participant lying recumbent. An aliquot for serum lipid analyses was stored at -25°C until thawed for the first time for the analyses. In study years 2001, 2007, and 2011 venous blood samples were drawn from the right antecubital vein of recumbent participants after a 12-hour fast and serum was separated, aliquoted and stored at -70°C until analysis. All lipid determinations were performed in duplicate in the same season (fall) and as simultaneously as possible.

1.3 Assessment of Lp(a) in YFS

The main aim of the present analyses was to examine the association between youth lipoprotein (a) (Lp(a)) level and adulthood cardiovascular events. In 1986, youth Lp(a) was measured in 2464 participants at ages 9-24 years by radioimmunoassay, according to the same principles as described for an immunoenzymatic assay using research kits from Pharmacia Diagnostics, Uppsala, Sweden. Serum samples pretreated by hydrolysis were incubated with two monoclonal antibodies directed toward different antigenic sites of apolipoprotein(a). One antibody is enzyme-labelled the other is attached to micro-Sepharose^{R}. The immunoreaction occurs in the wells of a Millititer^{TM} plate. These microtiter format plates are equipped with membranes that do not allow liquid to pass until suction applied. After completion of the immunoreaction free and bound tracer are separated by filtration across the membrane. The retained solid phase is washed and then contacted with chromogenic enzyme substrate. The enzymatic reaction is terminated by increasing the pH and the colored product is filtered across the membrane into a flat bottom micro plate. Absorbance was then measured in a photometric micro plate reader.

In subsequent follow-ups, Lp(a) levels have been also measured in adulthood, including in year 2001 (N=2281, ages 24-39 years), in year 2007 (N=2204, ages 35-45 years) and in year 2011 (N=2044, ages 39-49 years). Year 2001 plasma samples were stored at –70°C, and assessed in 2015 by using a sandwich ELISA method described by Kronenberg et al. (J. Lipid Res. 35, 1318–1323, 1994) with minor modifications. In brief, the ELISA plates (Nunc-Immuno MicroWell Maxi Sorp flat bottom design, MaxiSorp surface treatment; Thermo Fisher Scientific, Waltham) were coated using an affinity-purified polyclonal rabbit anti-human apolipoprotein(a) antibody in a final concentration of 5 µg/mL in 1×PBS containing 1 mg/mL NaN_3. The plates were incubated with 100 µL antibody dilution (3 hours, 37°C), washed 3× (1× PBS+0.05% v/v Tween-20), and blocked with 200-µL 0.1% wt/vol casein in 1× PBS pH 7.3 (30 minutes, 37°C). To ensure measuring each sample within the linear range of optical density, all samples were diluted into the ELISA plate twice (1:150 and 1:1500 in Assay Buffer [Microcoat, Bernried, DE] of 1:30 and 1:1000 predilutions in 1× PBS, pH 7.3). A 7-point standard curve

ranging from 0.32 mg/dL to 5 µg/dL was created (with an additional blank representing the zero point). Duplicate determinations of 4 reference samples were used as longitudinal interassay controls.

The coated plates were incubated with the analyte for 1 hour at 37°C. Detection was performed using a horseradish peroxidase-conjugated monoclonal antibody (1A2; in 0.1% wt/vol casein, 1× PBS, pH 7.3) directed against the KIV domain and not cross-reacting with plasminogen (1 hour, 37°C). After 3 washing steps as described above, 100 µL Blue Star TMB substrate (Adaltis, Guidonia Montecelio, IT) was added (30 minutes, room temperature). Reaction was stopped by adding 50 µL 0.5 mol/L sulfuric acid. Measurement of the absorption (dual wavelength, analyte: 450 nm, reference: 690 nm) was done using a Microplate Reader (BioRad Benchmark Plus; Bio-Rad Laboratories, Hercules), and concentrations were calculated based on the standard curve (expressed as mg/dL). All dilution and pipetting steps were done using liquid handling robotics (Tecan, Männedorf, CH).

In 2007 and 2011, Lp(a) was measured using an immunoturbidimetric method (Lp(a)-HA reagent, Wako Chemicals GmbH, Germany) on an AU400 instrument (Olympus, Japan). The assay utilizes an anti-human (goat) antibody that does not cross-react with plasminogen or apolipoprotein B (inter- and intra-assay coefficients of variation <3%). The assay utilizes turbidimetric immunoassay methodology in one step for quantitative determination of lipoprotein(a) in serum and plasma. When the sample is mixed with the buffer (Good's buffer, pH 7.5) and the antibody, Lp(a) in the sample combines specifically with anti-human lipoprotein(a) antibodies in the reagent to yield an insoluble aggregate that causes increased turbidity. The degree of turbidity can be measured optically and is proportional to the amount of Lp(a) protein mass in the sample. The within-run precision for this assay has less than 3 percent CV. The linear range of the assay is 1-100mg/dL.

1.4 Assessment of LDL-cholesterol in YFS

In the present analyses, the role of youth LDL-cholesterol in the association between Lp(a) and cardiovascular events. LDL-cholesterol was estimated using the Friedewald's equation from measured values of total cholesterol (TC), HDL-cholesterol and triglycerides).

In 1980, 1983, 1986, and 1989, TC concentrations were measured using a fully enzymatic CHOD-PAP method (Boehringer Mannheim, Mannheim, Germany) with OLLI 3000 and Kone CD analyzers (Kone Co., Espoo, Finland). In 1992, TC was measured using a fully enzymatic CHOD-PAP method (Merck, Darmstadt, Germany) with an automatic random access analyzer Olympus AU 510 (Olympus,Hamburg, Germany). Serum HDL cholesterol concentrations were measured from the supernatant after precipitation of very low density lipoprotein cholesterol and LDL cholesterol with dextran sulphate 500 000 (Pharmacia, Uppsala, Sweden). During 1980-1989, serum triglycerides were determined by using a fully enzymatic method (Boehringer Mannheim). In 1992, serum triglycerides were measured with colorimetric GPO-PAP methods (Merck) with an automatic analyzer Olympus AU 510 (Olympus).

In 2001, 2007 and 2011, TC levels were measured by the enzymatic cholesterol esterase–cholesterol oxidase method (Cholesterol reagent, Beckman Coulter Biomedical). The same reagent was used for estimating HDL-C levels after precipitation of very low density lipoprotein, intermediate-density lipoprotein, and low-density lipoprotein particles with dextran sulphate-MgCl2. Triglycerides concentration was determined using the enzymatic glycerol kinase–glycerol phosphate oxidase method (Triglyceride reagent, Beckman Coulter Biomedical, Ireland). All the above assays were performed on an AU400 instrument (Olympus, Japan).

1.5 Multiple Imputation in YFS

In the YFS, we imputed missing year 1986 Lp(a) values. We first imputed the Lp(a) values for individuals who had participated in later examinations and had at least one Lp(a) value available from study years 2001, 2007 and/or 2011. This subset included 74 cardiovascular cases and 3096 non-cases. The repeated Lp(a) measures were strongly correlated, the Spearman's rank order correlation coefficients varied between r=0.86-0.96. For example, the 25-year tracking correlation between year 1986 and 2011 was r=0.87, and the 4-year tracking correlation between 2007 and 2011 was r=0.96. Therefore, the imputed estimations of the missing year 1986 Lp(a) values in those individuals with at least one measured Lp(a) value can be considered highly reliable. Finally, we imputed the missing year 1986 Lp(a) values for all participants. The fully imputed data included 95 cardiovascular cases and 3484 non-cases.

Data were imputed using the Statistical Analyses System procedure MI (chained equations with fully conditional specification) and 20 replications of Cox Proportional Hazard Model regression to calculate relative risks and 95% confidence intervals to determine the associations between youth Lp(a) status and cardiovascular outcomes (using the procedure mianalyze and Rubin's rule for combining replications). We report the median of the p-values from the overall significance tests from the analyses on the imputed datasets.

We employed multiple imputation under the assumption of missing at random, utilizing existing data. The following variables were included in the imputation models:

Demographics: Age, sex, origin of birth (north-eastern vs. south-western Finland), area of residence in 2011 (5-class variable); Risk factor variables: Lp(a) measurements from study years 1986, 2001, 2007 and 2011; average values across age 6-18 of body mass index, LDL-cholesterol, non-HDL-cholesterol, systolic blood pressure, and smoking; adult values of body mass index, LDL-cholesterol, non-HDL-cholesterol, systolic blood pressure, and smoking; Genetic variables known to associate with Lp(a) levels, including variants of the LPA gene locus (rs3798220, rs783147, rs143431368, rs41272114 and rs10455872), and known modifying genes outside the LPA locus including the APOE236 genotype and the PCSK9 loss-of-function mutation R46L (rs11591147). The genotyping and related quality control was done by using Illumina 670K Custom array in Sanger Institute (UK).

1.6 Participant-specific curves by mixed model regression splines in YFS

To use all available repeatedly measured exposure data for the key risk variables that have been measured in the YFS in all study years, i.e. in 1980, 1983, 1986, 1989, 1992, 2001, 2007 and 2011, we have estimated participant-specific curves by mixed model regression splines. This enables the calculation of the area under the curve for continuous risk variables to evaluate the long-term burden of each measured attribute. These exposure variables have been calculated in separate models, prior to the modeling with cardiovascular events.

The covariance structure for the longitudinal setting was modelled by allowing for participant-specific regression spline coefficients, which were incorporated as random effects to the model. To avoid over-fitting, we reduced the number of knots (two knots on the calendar time from 1980 to 2011) for the participant-specific part from that of the fixed effects part (four knots on age from 3 to 34 years). To be more specific, we model the average and the participant-specific deviations from the average profile (and in-so-doing the covariance structure) by the mixed model regression spline

$$Y_{ij} = (\alpha + a_i) + (\beta + b_i)t_{ij} + \sum_{k=1}^{5} \gamma_k B_k(t_{ij}) + \sum_{l=1}^{3} c_{il} B_l(v_{ij}) + e_{ij}$$

where

- Yij is the log of risk variable of the ith participant at the jth occasion;
- tij is the age of the participant and vij is the (calendar time – 1980) at the jth measurement occasion;
- The B-functions are the mathematical basis functions for the spline construction.

Thus, the fixed effects part, $\alpha + \beta t_{ij} + \sum_{k=1}^{5} \gamma_k B_k(t_{ij})$, of the model estimates the average profile over the follow-up period. The role of the spline part is to allow departures from the linear mean trend in a flexible manner given by the data. In the final model, we allowed the intercept (α) to depend on the birth cohort and sex, and the slope (β) and the spline coefficients (γ) to be different for boys and girls.

The random-effects ai, bi and ci1-ci3, represent heterogeneity in trajectories across individuals, are assumed to be independent across participants, and normally distributed with zero mean and unstructured covariance matrix; finally, eij is the independent and normally distributed error term. The predicted values of the random effects for each participant, added to the estimated average trajectory, give the estimate of the individual trajectory (best linear unbiased prediction). The individual trajectories were then integrated to obtain area under the curve (AUC).

The mean profile was allowed to vary across birth cohorts and sex in terms of possibly different fixed effects parts. For body mass index and LDL-cholesterol, the AUC values were estimated for 3596 study participants using 12360 BMI values and 12552 LDL-cholesterol values. For systolic blood pressure, the AUC values were estimated for 3568 study participants using 8704 measurements. Between ages 6-18 BMI and LDL-cholesterol was measured on average 2.3 times/participant.

1.7 Bogalusa Heart Study (BHS)

Replication of the results was performed using data from the BHS, where Lp(a) measurements were completed during 1985-86 between ages 8 and 17 years. The BHS is a community-based, long-term investigation in Bogalusa, Louisiana, begun in 1973, focusing on the early natural history and risk factors for cardiovascular disease from childhood. A series of cross-sectional surveys was conducted, repeated every 2–3 years, allowing a longitudinal analysis in a cohort setting.

Height was measured to the nearest 0.1 cm using a manual height board, and weight was recorded to the nearest 0.1 kg using a balance beam metric scale. Blood pressures were recorded using a mercury sphygmomanometer. Serum lipid levels were measured according to the laboratory manual of the Lipid Research Clinics program. During 1973 to 1986, cholesterol and triglyceride levels were measured with a Technicon AutoAnalyzer II (Technicon Instrument Corp, Tarrytown, NY) according to the laboratory manual of the Lipid Research Clinics Program. Since 1987, these variables were determined by using an Abbott VP instrument (Abbott Laboratories, Abbott Park, Ill) by enzymatic procedures. Both chemical and enzymatic procedures met the performance requirements of the Lipid

Standardization Program of the Centers for Disease Control and Prevention, which routinely monitors the accuracy of measurements of total cholesterol, triglyceride, and HDL-C concentrations. Measurements on Centers for Disease Control and Prevention–assigned quality control samples showed no consistent bias over time within or between surveys. Serum lipoprotein cholesterols were analyzed by using a combination of heparin-calcium precipitation andagar-agarose gel electrophoresis procedures.

1.8 Carotid intima-media measurements in YFS and BHS

In YFS, carotid ultrasound studies were done in 2001 and 2007 follow-ups when participants were aged 24–45 years. Ultrasound studies were performed using Sequoia 512 ultrasound mainframes (Acuson, CA, USA) with 13.0 MHz linear array transducers. Left carotid artery was scanned by ultrasound technicians following a standardized protocol. The image was focused on the posterior (far) wall, and images were recorded from the angle showing the greatest distance between the lumen-intima interface and the media-adventitia interface. A scan including the beginning of the carotid bifurcation and the common carotid artery was recorded and stored in digital format on optical discs for subsequent off-line analysis. All scans were analyzed by one reader blinded to participants' details. The best quality end-diastolic frame was selected. Several measurements of the common carotid far wall were taken approximately 10 mm proximal to the bifurcation to derive maximal carotid IMT. The mean of maximum values done in study years 2001 and 2007 were used in the analyses. Digitally stored images were scanned for the existence of carotid atherosclerotic plaques, defined as a focal structure of the arterial wall protruding into the lumen > 50% compared to the adjacent intima-media thickness. All plaques were located in the carotid bifurcation. To assess reproducibility of IMT measurements, we re-examined 60 participants 3 months after the initial visit (2.5% random sample). The between-visit coefficient of variation of IMT measurements was 6.4%. To assess reproducibility of IMT image analysis, 113 scans were re-analyzed by a second observer. The between observer coefficient of variation was 5.2%.

In BHS, carotid ultrasound studies were as part of the surveys in done in study years between 2001-2010. Trained sonographers blinded to risk factor data performed ultrasound examinations with a Toshiba Sonolayer SSH160A (Toshiba Medical, Tokyo, Japan), a 7.5-MHz linear array transducer, on participants in the supine position with the head slightly extended and turned to the opposite direction of the carotid artery being studied. Images were recorded at the common carotid, carotid bulb (bifurcation), and internal carotid arteries bilaterally according to previously developed protocols for the Atherosclerosis Risk in Communities Study. Images were recorded on S-VHS tapes and read by certified readers from the Division of Vascular Ultrasound Research by using a semiautomatic ultrasound image processing program developed by the California Institute of Technology Jet Propulsion Laboratory (Pasadena). The mean of the maximum carotid IMT readings of 3 right and 3 left far walls for common, bulb, and internal segments was used. The mean of maximum values done in study years 2001-2010 were used in the analyses.

Supplementary Table 1. Characteristics of the participants in the analyses sample*.

	YFS*	BHS*
Age, years†		
Mean	15.7	13.8
Range	9-24	8-21
Males, %	48	41
Body mass index, kg/m2, mean (SD)	20.0 (3.4)	20.4 (4.6)
Regular smoking in adolescence, %	27	47
Parental education, %		
-less than high-school	50.2	7.9
-high-school	12.8	35.3
-high-school but no college degree	25.1	24.5
-college degree	11.9	32.2
LDL cholesterol, mmol/L, mean (SD)	2.97 (0.82)	2.44 (0.60)
Lp(a) mg/dL, mean (median)		
-baseline: YFS 1986, BHS 1984	11.4 (5.3)	19.2 (10.4)
-2001 (ages 24-39)	12.7 (6.3)	n/a
-2007 (ages 30-45)	14.1 (7.3)	n/a
-2011 (ages 34-49)	14.7 (8.1)	n/a
High Lp(a) (>30 mg/dL), %		
-1986 (baseline)	10.1	25.5
-2001 (ages 24-39)	12.8	n/a
-2007 (ages 30-45)	14.2	n/a
-2011 (ages 34-49)	14.7	n/a

*YFS and BHS (White individuals) with data on youth Lp(a) and adult cardiovascular events

†Age at the time of youth Lp(a) measurement, year 1986 in YFS and 1984 in BHS.

n/a = not available

Supplementary Table 2. Number of cardiovascular diagnoses in the analyses sample* stratified by youth Lp(a) level using cut-point of 30 mg/dL.

	Lp(a)<30 mg/dL	**Lp(a)≥30 mg/dL**
YFS		
Total N	2208	247
Any CVD outcome (N, %)	38 (1.7%)	8 (3.2%)
Coronary artery disease (N, %)	24 (1.1%)	6 (2.4%)
Ischemic stroke (N, %)	8 (0.4%)	2 (0.8%)
Peripheral artery disease (N, %)	2 (0.4%)	-
Transient ischemic attack or temporary stroke (N, %)	5 (0.2%)	1 (0.4%)
Blocked carotid artery (N, %)	-	-
Abdominal aneurysm (N, %)	1 (0.05%)	-
BHS		
Total N	437	150
Any CVD outcome (N, %)	8 (1.8%)	7 (4.7%)
Coronary artery disease (N, %)	8 (1.8%)	7 (4.7%)
Ischemic stroke (N, %)	-	1 (0.7%)
Peripheral artery disease (N, %)	-	1 (0.7%)
Transient ischemic attack or temporary stroke (N, %)	-	1 (0.7%)
Blocked carotid artery (N, %)	-	1 (0.7%)
Abdominal aneurysm (N, %)	-	1 (0.7%)

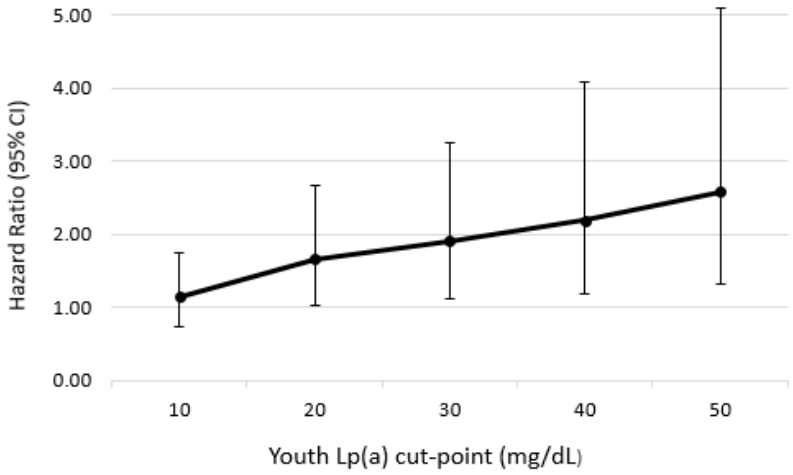
In YFS, there were 46 individuals with one or more cardiovascular diagnoses: 43 individuals with one diagnoses and 3 individuals with two separate diagnoses. In BHS, there were 15 individuals with one or more cardiovascular diagnoses. All of them had been diagnosed with coronary artery disease, and one individual had additionally multiple diagnoses.

* YFS and BHS (White individuals) with data on youth Lp(a) and adult cardiovascular events.

2. Extended Results

2.1 Assessing non-linearity

We also explored the possibility of a nonlinear relation between Lp(a) and cardiovascular outcomes. This was done by two approaches. First, we ran the Cox-proportional hazard model with fully imputed data by including the second-order interaction term for loglpa*loglpa. The p-value (p=0.48) did not provide support for non-linear relationship. Second, we calculated the Hazard Ratios in the fully imputed data by using different cut-points to define high-risk youth Lp(a) level. High Lp(a) was defined by cut-points 10, 20, 30, 40 and 50 mg/dL. The risk of cardiovascular events associated with each cut-point is shown in the Figure below. There was no evidence of non-linearity in the association.



Supplementary Figure 1. Age and sex adjusted risk of cardiovascular events in mid-adulthood in relation to different cut-points of youth Lp(a) level in the Young Finns Study (fully imputed data).

2.2 Assessing the effect of high LDL-cholesterol levels on the association between Lp(a) and cardiovascular events

To address how the association of Lp(a) and cardiovascular events changes after excluding individuals with extreme LDL-cholesterol levels, we re-analyzed the data by eliminating individuals with high youth LDL-cholesterol values, who by based on their phenotype could be e.g. potential patients with familiar hypercholesterolemia.

The results are shown in a Table below:

Supplementary Table 3. Effect of excluding individuals with high youth LDL-cholesterol levels on the association between youth Lp(a) and adult cardiovascular events in the Young Finns Study (fully imputed data).

Youth LDL cholesterol	Cases	Non-cases	HR	95% CI
No exclusion	95	3484	1.96	1.35-2.57
LDL-c ≤ 5.5 mmol/L (≤ 213 mg/dL)	95	3457	1.97	1.36-2.58
LDL-c ≤ 5.0 mmol/L (≤ 193 mg/dL)	92	3450	1.76	1.13-2.39
LDL-c ≤ 4.5 mmol/L (≤ 174 mg/dL)	88	3303	1.79	1.13-2.45
LDL-c ≤ 4.0 mmol/L (≤ 155 mg/dL)	68	2891	1.90	1.11-2.68
LDL-c ≤ 3.5 mmol/L (≤ 135 mg/dL)	38	2094	1.88	0.81-2.96

The risk estimates associated with high Lp(a) levels did not materially change after excluding individuals with elevated youth LDL-cholesterol levels by using different cut-points.

Therefore, the data suggest that the association between Lp(a) and cardiovascular outcomes is observed after restricting the analyses to population without LDL-cholesterol dyslipidemia.