Association between Lp(a) and T2D: A Mendelian Randomization Study

Keywords
lipoprotein(a), diabetes, Mendelian randomization

Abstract

Introduction
Blood lipoprotein(a) (Lp(a)) levels have been observed to be inversely correlated with type 2 diabetes (T2D). In this Mendelian Randomization (MR) study, the causal impact of genetically-predicted Lp(a) on T2D was assessed.

Material and methods
A two-sample MR analysis was conducted. Data were obtained from UKBiobank and FinnGen consortia. Primary analysis was based on inverse-variance-weighted mean (IVM) approach.

Results
No statistically significant association between the genetically predicted levels of Lp(a) and T2D was detected (p=0.362) in IVM analysis involving data of 563,420 patients.

Conclusions
Genetically predicted Lp(a) concentration does not appear to causally influence the risk of T2D.

Association between Lp(a) and T2D: A Mendelian Randomization Study

[Author names]

Telephone: [Telephone]

*Email: [Email]

1. [Affiliations]

Abstract:

Introduction:

Blood lipoprotein(a) (Lp(a)) levels have been observed to be inversely correlated with type 2 diabetes (T2D). In this Mendelian Randomization (MR) study, the causal impact of genetically-predicted Lp(a) on T2D was assessed.

Methods:

A two-sample MR analysis was conducted. Data were obtained from UKBiobank and FinnGen consortia. Primary analysis was based on inverse-variance-weighted mean (IVM) approach.

Results:

No statistically significant association between the genetically predicted levels of Lp(a) and T2D was detected (p=0.362) in IVM analysis involving data of 563,420 patients.

Conclusions:

Genetically predicted Lp(a) concentration does not appear to causally influence the risk of T2D.

Keywords: Mendelian randomization, lipoprotein(a), diabetes

Introduction

Lipoprotein(a) [Lp(a)] is a unique human plasma lipoprotein. It consists of an LDL-like core and covalently bound apolipoprotein(a)[1]. Apolipoprotein(a) [Apo(a)] is encoded by the *LPA* gene and is highly homologous to plasminogen in its protein sequence[1]. Lp(a) has a highly variable concentration between individuals, which is in large part defined by single nucleotide polymorphisms (SNPs) and copy number variants (CNV) found in the *LPA* gene locus[1,2].

Despite over 60 years of research, key functions of Lp(a) still remain enigmatic. On the other hand, Lp(a) has a clearly detrimental role in human disease since high Lp(a) levels are correlated with coronary artery disease and myocardial infarction[2]. These are thought to occur due to the prothrombotic action of Lp(a) and especially through promotion of atherosclerosis[3].

Although high levels of Lp(a) are known to cause cardiovascular disease[2], low Lp(a) has been epidemiologically shown to correlate with the incidence of type 2 diabetes (T2D)[4]. In the case of T2D, the mechanism behind this association appears to be much less well elucidated[5].

Because Lp(a) levels are in large parts determined by genetics, they seem particularly amenable to Mendelian randomization (MR). To further elucidate the connection between T2D and Lp(a) levels, we have performed a MR analysis to analyze the impact of genetically predicted Lp(a) concentration on the incidence of T2D.

Methods

Study design

In this research, we conducted a two-sample MR investigation, utilizing Genome-Wide Association Studies (GWAS) data sourced from publicly accessible repositories. MR is an epidemiological approach that relies on instrumental variable analysis. It involves the use of genetic variants, commonly single-nucleotide polymorphisms (SNPs), linked to a

modifiable risk factor like high blood pressure to make inferences about causality. By leveraging these genetic markers as proxy indicators, MR helps minimize bias arising from confounding factors since alleles are randomly inherited at conception. Moreover, it also avoids bias due to reverse causation, as the occurrence of a disease cannot influence an individual's genotype [6].

Data sources

Exposure: We extracted SNP as instrumental variables (IVs) associated with Lp(a) concentration from genome-wide association study (GWAS) data downloaded from Neale lab (UK Biobank). Because the measured Lp(a) concentration was positively skewed, inverse rank-normalized data were used. European ancestry female and male individuals were included in this research (*N* = 361,194) [7].

Outcome: Summary-level data for T2D were obtained from Finnish FinnGen consortium [8]. We used data from the eighth version of the database, which included 17,268 cases and 184,778 controls, with an analysis covering 16,380,418 variants.

MR analyses: $P<5x10^{-8}$ was accepted as a genome-wide significance threshold. To reduce the risk of any potential weak instrument bias, F-statistics were calculated based on the formula F=(beta/se)^2 [6]. Only the SNPs with F-statistics>10 were considered potential IVs. To ensure the independence of IVs, SNP in the linkage of disequilibrium (LD) were excluded. The TwoSampleMR R package was used to clamp data with a threshold r2<0.001 [9]. If selected SNPs were unavailable in the outcome dataset, they were replaced with proxies in LD of r2 > 0.8 or excluded from further MR analysis. Later, for selected SNPs, the potential association with confounding or risk factors for T2D were evaluated using PhenoScanner V2[10]. Next, variant harmonization was conducted using the TwoSampleMR package between datasets to confirm that the association between SNPs and exposure and between SNPs and the outcome reflected the same allele. To further validate our results, aside from the aforementioned method of SNP selection, we

also analyzed variants listed by ESC guidelines concerning Lp(a) including a subset of loss of function variants[11].

As the main analysis for evaluating the causal effect estimates in our study, we used inverse variance weighted (IVW) method[12]. Sensitivity analyses were performed using MR-Egger, weighted median, and weighted mode approaches [13–15].

Additionally, to detect the existence of heterogeneity, horizontal pleiotropy, and outlier SNPs we performed several statistical tests. Cochran's Q test was used to measure the heterogeneity between variant-specific causal estimates (IVW and MR-Egger regression)[16]. MR-Egger intercept was calculated to measure the presence of horizontal pleiotropy[17]. To detect the potential presence of pleiotropic outlier SNPs, the MR-pleiotropy residual sum and outlier (MR-PRESSO) tests were applied[17]. Finally, the leave-one-out analyses were performed to check whether any of the analyzed SNPs are strongly associated with the exposure, which may dominate the estimate of the causal effect.

Statistical analysis

All statistical data analyses were conducted with R software (version 4.1.1) using the "TwoSampleMR" (version 0.5.6), and the „MR-PRESSO" package (version 1.0). A p-value <0.05 was considered statistically significant for all tests.

Ethics statement

Ethical approval for this study was not required as our analyses were based on summary statistics from published GWAS, or the data were publicly accessible and no individual-level data were used.

Results

According to the accepted criteria and after searching in Phenoscanner, we finally identified as IVs for Lp(a) 23 SNPs which were used in MR analysis. Summary characteristics of the final IVs for Lp(a) and T2D are listed in **Supplementary Table 1**.

Figure 1

The results of estimating the causal effect between Lp(a) and T2D are presented in **Figure 1A**. The main MR analysis showed that the concentrations of genetically predicted Lp(a) (OR=1.008; 95% CI=0.991-1.026; P=0.362) were not associated with T2D. This lack of association was supported by MR Egger (OR=1.016; 95% CI=0.989-1.043; P= 0.256), Weighted median (OR=1.015; 95% CI=0.998-1.03; P= 0.098), and Weighted mode (OR=1.012; 95% CI=0.995-1.029; P=0.189) approach in sensitivity analysis.

As far as other analyses are concerned, we detected weak heterogeneity between Lp(a) and T2D SNPs using Cochran's Q test (P_{IVW} = 0.03). There was no evidence of horizontal pleiotropy across the analyses in the MR-Egger regression ($P_{intercept}$ = 0.46). No outlier SNPs were observed in the MR-PRESSO analysis (P=0.06).

In the leave-one-out analysis, there was no significant change in the risk estimations for genetically predicted Lp(a) levels and preeclampsia risk after removing 1 SNP at a time, demonstrating that the causal association was not driven by any specific SNPs. Only 1 SNP (rs118039278) relatively affected the robustness of the results (**Figure 1B**). The analysis of variants included in EAS recommendations[11] did not yield a statistically significant result (OR=1.003; 95% CI=0.977-1.031; P=0.781 for IVM). No statistically significant impact of Lp(a) levels on T2D as determined by loss-of-function variants was noted (OR=0.913; 95% CI=0.807-1.032; P=0.145 for IVM).

Discussion

MR has not broadly supported the causal role of Lp(a) in T2D[5]. This two-sample MR analysis utilizing data of a total of 563,420 patients is another one such instance. There are several possible reasons behind this.

There is some evidence that insulin reduces apo(a) expression in hepatocytes *in vitro* – this further complicates the issue of Lp(a)-T2D connection, hinting at possible reverse causation[18] that nevertheless does not influence our MR analysis. So far, only one, one-sample MR study examining the relationship between serum insulin and Lp(a) has been published and it did not demonstrate a significant reverse-causation[19].

The molecular mass of apo(a) is highly variable between individuals due to copy number variants (CNV) of Kringle-IV Type 2 (KIV-2) domains in the LPA gene[1]. This complicates the potential role of Lp(a) in T2D, since some reports emphasize the role of apo(a) KIV-2 CNVs as the potential cause of T2D rather than merely low Lp(a) concentration[5].

Moreover, some have suggested that the relationship between Lp(a) and T2D is non-linear and only the lowest Lp(a) concentrations increase the risks appreciably[5]. This means that standard MR approaches like the one used here may not yield reliable results, as linear relationship between exposure and outcome is assumed[6]. Nevertheless, even the analysis of loss-of-function variants did not show Lp(a) to causally influence T2D risk, partially negating this concern.

In conclusion, this MR study did not show the correlation between Lp(a) and T2D to be causal. Further studies, including those utilizing animals, *in vitro* models and human patients, and epidemiological data are needed to elucidate a clear mechanism behind the observed Lp(a)-T2D correlation. This is of particular importance due to an on-going research on specific Lp(a)-targeting therapies and an already known Lp(a)-lowering

effect of PCSK9 inhibitors[20]. Since that time one is strongly recommended to reduce Lp(a) levels with the available methods, which is a significant residual CVD risk factor[2].

References:

1. Cybulska B, Kłosiewicz-Latoszek L, Penson PE, [Author name]. What do we know about the role of lipoprotein(a) in atherogenesis 57 years after its discovery? Prog Cardiovasc Dis. 2020 May-Jun;63(3):219-227.

2. Sosnowska B, Stepinska J, Mitkowski P, Bielecka-Dabrowa A, Bobrowska B, Budzianowski J, Burchardt P, Chlebus K, Dobrowolski P, Gasior M, Jankowski P, Kubica J, Mickiewicz A, Mysliwiec M, [Author name], Prejbisz A, Rajtar-Salwa R, Wita K, Witkowski A, Gil R, [Author name]. Recommendations of the Experts of the Polish Cardiac Society (PCS) and the [Author's institution] on the diagnosis and management of elevated lipoprotein(a) levels. Arch Med Sci. 2024 Jan 31;20(1):8-27.

3. Boffa MB, Koschinsky ML. Lipoprotein (a): truly a direct prothrombotic factor in cardiovascular disease? *Journal of Lipid Research*. 2016;57(5):745-757. doi:10.1194/jlr.R060582

4. Skoumas I, Andrikou I, Grigoriou K, et al. Lipoprotein(a), metabolic profile and new-onset type 2 diabetes in patients with familial combined hyperlipidemia: A 9 year follow-up study. *Journal of Clinical Lipidology*. 2023;17(4):512-518. doi:10.1016/j.jacl.2023.05.103

5. Lamina C, Ward NC. Lipoprotein (a) and diabetes mellitus. *Atherosclerosis*. 2022;349:63-71. doi:10.1016/j.atherosclerosis.2022.04.016

6. Burgess S, Thompson SG. *Mendelian Randomization: Methods for Causal Inference Using Genetic Variants*. Second edition. CRC Press; 2021.

7. GWAS of UK Biobank biomarker measurements. Neale lab. Published September 16, 2019. Accessed March 30, 2024. http://www.nealelab.is/blog/2019/9/16/biomarkers-gwas-results

8. Kurki MI, Karjalainen J, Palta P, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature*. 2023;613(7944):508-518. doi:10.1038/s41586-022-05473-8

9. Phenoscanner. Phenoscanner. http://www.phenoscanner.medschl.cam.ac.uk/about/

10. Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Statistics in Medicine*. 2017;36(11):1783-1802. doi:10.1002/sim.7221

11. Kronenberg F, Mora S, Stroes ESG, et al. Lipoprotein(a) in atherosclerotic cardiovascular disease and aortic stenosis: a European Atherosclerosis Society consensus statement. *European Heart Journal*. 2022;43(39):3925-3946. doi:10.1093/eurheartj/ehac361

12. Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. *Epidemiology*. 2017;28(1):30-42. doi:10.1097/EDE.0000000000000559

13. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genetic Epidemiology*. 2016;40(4):304-314. doi:10.1002/gepi.21965

14. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *International Journal of Epidemiology*. 2017;46(6):1985-1998. doi:10.1093/ije/dyx102

15. Greco M FD, Minelli C, Sheehan NA, Thompson JR. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Stat Med*. 2015;34(21):2926-2940. doi:10.1002/sim.6522

16. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015;44(2):512-525. doi:10.1093/ije/dyv080

17. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet*. 2018;50(5):693-698. doi:10.1038/s41588-018-0099-7

18. Neele DM, de Wit EC, Princen HM. Insulin suppresses apolipoprotein(a) synthesis by primary cultures of cynomolgus monkey hepatocytes. *Diabetologia*. 1999;42(1):41-44. doi:10.1007/s001250051110

19. Buchmann N, Scholz M, Lill CM, et al. Association between lipoprotein(a) level and type 2 diabetes: no evidence for a causal role of lipoprotein(a) and insulin. *Acta Diabetol*. 2017;54(11):1031-1038. doi:10.1007/s00592-017-1036-4

20. Sosnowska B, Surma S, [Author name]. Targeted Treatment against Lipoprotein (a): The Coming Breakthrough in Lipid Lowering Therapy. Pharmaceuticals (Basel). 2022 Dec 16;15(12):1573.

rs	chr	pos	ref	alt	beta Lp(a)	se Lp(a)	mlog pval Lp	af_alt UKBi
rs112376176	6	160874032	C	T	-0,406	0,00695	-1709	0,04378
rs112842440	6	160955713	G	T	-0,03014	0,00946	-6,537	0,0235
rs117945468	6	161001215	G	C	0,1924	0,00924	-220,1	0,02432
rs118039278	6	160985526	G	A	1,805	0,00569	-50350	0,07135
rs12214416	6	160910517	T	A	-0,3578	0,00717	-1249	0,04127
rs143053678	6	160934405	G	A	-0,5233	0,00977	-1440	0,02479
rs144177163	6	161164290	C	A	-0,3597	0,00801	-1011	0,03401
rs145989243	6	160969113	G	A	-0,3646	0,02578	-102,9	0,003144
rs147555597	6	160911596	G	A	1,651	0,01586	-5426	0,008402
rs187614196	6	161001087	C	T	0,6891	0,04973	-98,85	0,000954
rs191243877	6	160979083	C	T	-0,3142	0,02688	-70,98	0,003276
rs191592502	6	161172374	C	T	-0,2625	0,01218	-235,3	0,01509
rs192559810	6	161150306	G	A	-0,4251	0,02406	-159,1	0,00406
rs2504925	6	160876394	T	C	0,185	0,00348	-1417	0,7855
rs41266385	6	160952333	A	C	0,3124	0,00611	-1310	0,05801
rs41269133	6	161087863	T	C	-0,3562	0,00523	-2324	0,08222
rs41271036	6	161015301	A	G	-0,4953	0,01294	-736,5	0,01267
rs4252117	6	161143376	A	G	-0,2806	0,00317	-3932	0,2841
rs73596816	6	161017363	G	A	0,9843	0,00776	-8059	0,03513
rs80145669	6	161022495	G	A	0,2541	0,00802	-506,2	0,03265
rs9456544	6	160908853	C	T	-0,247	0,0098	-321,2	0,022
rs9456577	6	161145246	A	C	0,3731	0,00876	-910	0,02765
rs9457927	6	160910282	A	G	0,9146	0,01597	-1645	0,008068

f_statistic Lp(a)	nearest_genes	mlog pval T	beta T2D	se T2D	af_alt FinnGe
3409,6327855938200	SLC22A3	0,251838	0,0163016	0,0279668	0,025168
10,14445724	LPA	0,584695	0,0376512	0,0334404	0,0171908
433,7647136	LPA	0,292586	-0,0259314	0,0393423	0,0124858
100701,3371	LPA	0,735174	0,0273673	0,0205997	0,0465967
2489,552146	SLC22A3	0,190868	0,00938189	0,0203245	0,0486648
2871,818403	LPA	0,806754	-0,0419765	0,0295921	0,0235567
2014,569244	PLG	0,089654	-0,005654	0,0239635	0,0346768
200,0172733	LPA	0,352296	0,0544547	0,0711916	0,00369478
10836,46869	SLC22A3	0,298996	0,0302654	0,0451183	0,00939822
192,0116508	LPA	0,01159	0,00672041	0,203582	0,00072124
136,6324427	LPA	0,010428	-0,0024028	0,0807959	0,00301514
464,4768133	PLG	0,182425	-0,0167262	0,0376683	0,0142652
312,1698559	PLG	0,367556	0,104909	0,132641	0,00109446
2826,083366	SLC22A3	1,26083	0,0219048	0,0114083	0,823528
2610,785438	LPA	0,019689	-0,0013704	0,0246562	0,0338923
4638,573606	LPA	0,828888	0,0178763	0,012366	0,146951
1465,102817	LPA	0,969848	0,0375441	0,0233055	0,0368587
7855,13245	PLG	0,242169	0,00553688	0,0098126	0,26862
16109,84515	LPA	0,112219	-0,010368	0,0358283	0,0156553
1004,833055	LPA	0,29449	0,0133231	0,0201073	0,0498384
635,5050943	SLC22A3	1,88211	0,055288	0,0222888	0,0395938
1812,365391	PLG	0,293686	0,0171503	0,0259408	0,0296669
3279,8392728382000	SLC22A3	0,294975	-0,0279519	0,0421289	0,0109733

af_alt_cases F	af_alt_controls FinnGen
0,0250175	0,0251855
0,0177513	0,0171258
0,012283	0,0125093
0,0485746	0,0463674
0,0488929	0,0486383
0,0229386	0,0236283
0,0344645	0,0347014
0,00403847	0,00365493
0,00965664	0,00936825
0,00071849	0,000721553
0,00295123	0,00302255
0,0141997	0,0142728
0,00120034	0,00108218
0,825626	0,823285
0,0336252	0,0339233
0,148169	0,14681
0,037538	0,03678
0,269652	0,268501
0,0156541	0,0156555
0,0501006	0,0498079
0,0414663	0,0393767
0,0300355	0,0296242
0,0105707	0,01102

rs	rsid of SNP
chr	Chromosome of SNP
pos	Position on a chromosome based on GRCh37
ref	Reference allele
alt	Alternative allele
beta Lp(a)	Beta for correlation between alternative allele and Lp(a) levels
se Lp(a)	Standard error for correlation between alternative allele and Lp(a) levels
pval Lp(a)	Negative log of p-value for correlation between alternative allele and Lp(a) leve
af_alt UKB	Alternative allele frequency in exposure data
f_statistic l	F statistic for exposure data
nearest_ge	Nearest gene
pval T2D	Negative log of p-value for correlation between alternative allele and T2D
beta T2D	Beta for correlation between alternative allele and T2D
se T2D	Standard error for correlation between alternative allele and T2D
af_alt Finn	Alternative allele frequency in outcome data
af_alt_cas	Alternative allele frequency in outcome data (cases only)
af_alt_con	Alternative allele frequency in outcome data (controls only)

Is

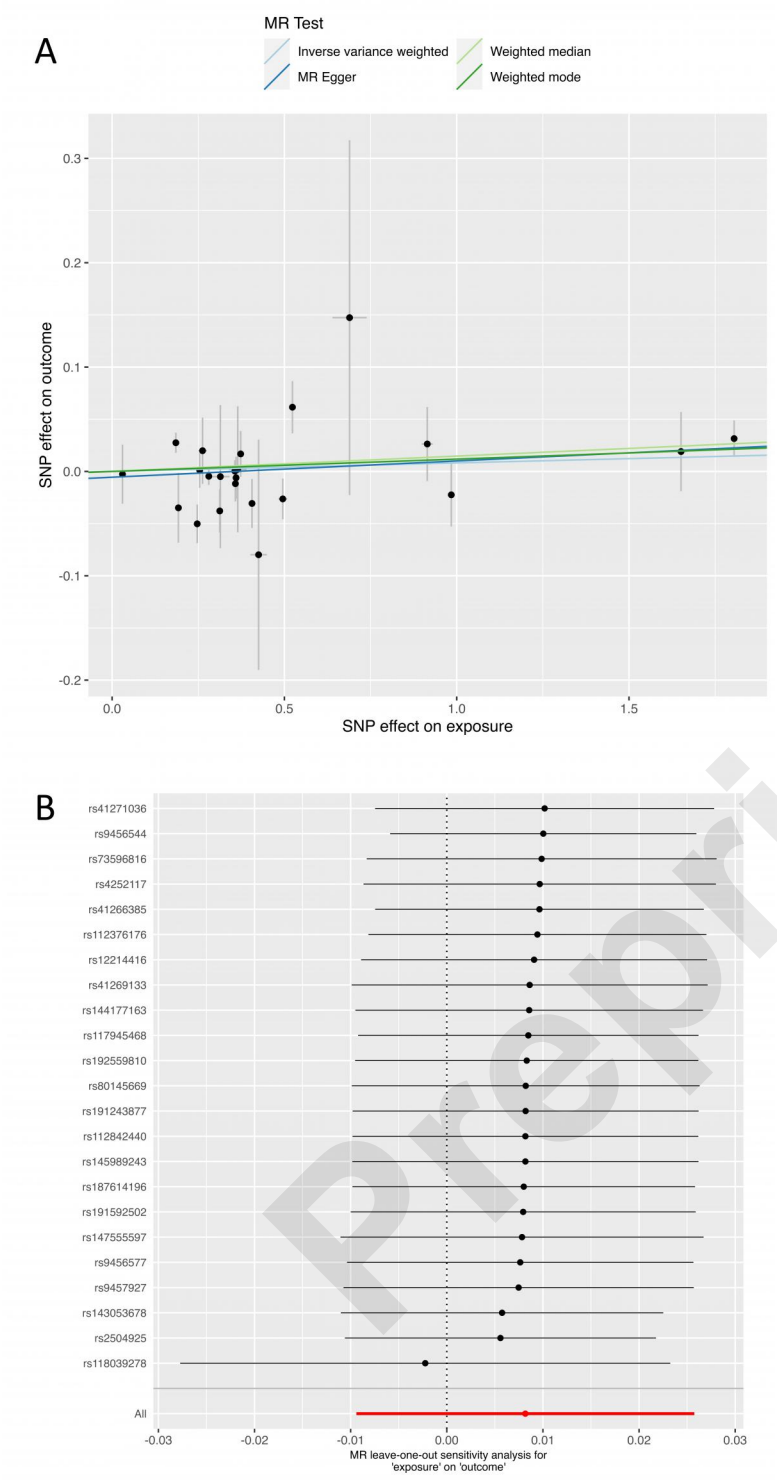

Figure 1. A) Genetic associations between the effect alleles of SNPs, exposure, and outcome. X axis - effect of SNP on exposure (Type 2 Diabetes). Y axis - effect of SNP on outcome (lipoprotein(a)). Each dot represents SNP + standard error. Lines represent Mendelian randomization estimates of different tests.
B) Leave-one-out analysis. Each dot represents a single inverse-variance-weighted mean estimate computed by leaving out the variant specified on the right. Lines represent 95% confidence intervals.