

行政院國家科學委員會補助專題研究計畫成果報告

Familial Correlation and Segregation Analysis of HDL cholesterol in Adults in Taiwanese: Based Health Examination Population 成年人高密度脂蛋白膽固醇濃度的家族相關與分離分析研究: 以健檢病人為指標個案之家族遺傳研究設計

計畫類別： 個別型計畫 整合型計畫

計畫編號：NSC - 89 - 2314 - B - 002 - 403

執行期間：89年8月1日至90年7月30日

計畫主持人：簡國龍

本成果報告包括以下應繳交之附件：

赴國外出差或研習心得報告一份

赴大陸地區出差或研習心得報告一份

出席國際學術會議心得報告及發表之論文各一份

國際合作研究計畫國外研究報告書一份

執行單位：國立台灣大學醫學院

中 華 民 國 90年8月15日

行政院國家科學委員會專題研究計畫成果報告

成年人高密度脂蛋白膽固醇濃度的家族相關與分離分析研究：以健檢病人為指標個案之家族遺傳研究設計

Familial Correlation and Segregation Analysis of HDL cholesterol in Adults in Taiwanese: Based on Health Examination Population

計畫編號：NSC - 89 - 2314 - B - 002 - 403

執行期限：89年8月1日至90年7月31日

主持人：簡國龍 執行機構及單位名稱：國立台灣大學醫學院

一、中文摘要

由大規模世代研究及臨床研究結果得知血脂異常是心血管疾病重要的危險因子，而其中以高密度脂蛋白膽固醇(HDL-C)是重要的保護因子，國外雙胞胎研究及家族研究均顯示血脂異常有高度的家族聚集傾向。本研究設計經由三級醫院健檢其一等親家人作為研究對象，收集家族親屬包括一等親(兄弟姐妹、父母)血脂值進行 HDL-C 濃度的家族研究。自 1998 年 8 月開始到 1999 年 9 月收集總共 1313 位樣本，其中包括有 530 對配偶關係，76 對親子及以 52 對手足之間的配對。結果顯示：各種血脂質如 cholesterol、triglyceride、HDL-C 及 LDL-C 值，均可看到在親子之間的相關係數均是很高，而且以 HDL-C 的 $r=0.436$ 最高。在手足之間的相關係數則是以 HDL-C 最高，高達 0.505。混合分析結果顯示，以 3 個成份的混合分析可以最佳解釋模型，因此我們認為 HDL-C 的分佈由 3 個成份的混合模型為最佳模型。其中標準化平均值分別為 -0.346、0.833、4.370，而變異數為 0.516、1.043 及 0.085，而分別所作的比率為 71.5%、28.2% 及 0.3%。進一步在分離分析時發現隱性基因為一較好的模型。我們確實可以發現有一隱性基因作用在 HDL-c 濃度的控制。其中其基因頻率為 0.280，三個平均值為 79.5，50.5 及 50.5 mg/dL，而族群變異數為 127.1。配偶之間的相關係數為 0.285，而子代之間的相關係數為 0.466，估計其遺傳率為 62.7%。因此，我們結論：在以醫院健檢的病人之家族成員中，HDL-C 有明顯的遺傳因素存在，而且是一隱性基因的作用。

關鍵詞：高密度脂蛋白膽固醇、家族相關分析、混合分析、分離分析、家族研究

Abstract

Introduction: There were enough evidences that there was strong familial aggregation of HDL-C concentration in Caucasian, and the estimated heritability was around 20-40% from twin and pedigree studies. But there were insufficient evidences of genetic components among adult Chinese.

Materials and Methods: We collected 1,313 subjects, including 530 spouse, 76 parent-offspring and 52 sibling pairs from Aug, 1998 to Sep, 1999. HDL-C was measured following precipitation of apolipoprotein B-containing lipoproteins with phosphotungstic acid and magnesium ions (Boehringer Mannheim, Germany). Residual HDL-C levels, adjusted after age and gender, are used to further

genetic analysis, including familial correlation, commingling analysis, and complex segregation analysis.

Results and Conclusion: There are significantly high correlation coefficients in parent-offspring and sibling pairs in HDL concentration (0.436 and 0.505, respectively). Commingling analyses demonstrated that three-component distribution was the best-fit model to explain the variation of HDL, with standardized mean values with -0.346, 0.833, 4.370; with proportions of 71.5%, 28.2% and 0.3%. Complex segregation analyses clearly showed that the mode of inheritance of HDL-C was controlled by one recessive major gene effect, with gene frequency of 0.280, and three means of 79.5, 50.5 and 50.5 mg/dL. The variance was 127.1, and correlation coefficients in spouse and siblings were 0.285 and 0.466, respectively, and the estimated heritability was 62.7%. So we concluded that in the population based on hospital health examination, there are significant genetic components in HDL-C, and one recessive major gene effect could be clearly demonstrated.

Keywords: High density lipoprotein cholesterol, family correlation, commingling analysis, complex segregation analysis,

二、緣由與目的

高密度脂蛋白膽固醇(HDL-C)在心血管疾病的發生為一重要的保護因子。由大規模世代研究及臨床試驗結果證實，HDL-C 濃度與心血管疾病有負相關的關係。一些外在的環境因子與 HDL-C 的關係也較模糊，因此學者研究興趣均擺在 HDL-C 遺傳成份上。文獻回顧顯示 HDL-C 大多數的研究顯示為單一基因效應，而少數只証實為多重基因背景。本研究目的是以台大醫院健康檢查受檢者作為指標個案，成員了解 HDL-C 濃度的分佈狀況，並控制年齡、性別的因素研究估計其遺傳率。利用家族相關研究，了解 HDL-C 在家族不同配對之間的相關性研究。利用混合分析模型，了解族群中 HDL-C 分佈最佳的預測。利用分離分析方法，研究 HDL-C 其遺傳模式為何。探討 HDL-C 的基因遺傳模式為單一基因或多重基因控制，其家族成員如配偶、親子及手足之間配對的相關係數為何。

三、結果與討論

自 1998 年 8 月開始到 1999 年 9 月收集總共健檢 4331 位成年人，其中排除單一受檢者，以配偶及有一等家族關係者納入本研究，一共收集到 1313 位樣本，其中包括有 530 對配偶關係，76 對親子及以 52 對手足

之間的配對。Table 1 顯示在研究族群中一些危險因子分佈的情形，包括其平均值、標準差、skewness 及 kurtosis 值。我們發現大部分的血脂質，其 skewness 及 kurtosis 值均接近 0。表示其分佈大致為一常態分佈。各種血脂質如 cholesterol triglyceride HDL-C 及 LDL-C 或是 TG/HDL 對 ratio 值，均可看到在親子之間的相關係數均是很高(Table 2)，而且以 HDL-C 的 $r=0.436$ 最高。而在配偶之間的相關係數變化比較大，膽固醇及 LDL-C 較低，為 0.054 及 0.074 而已，為 HDL 及 $\log(\text{TG}/\text{HDL})$ 則為 0.201 及 0.142，而在手足之間的相關係數則是以 HDL-C 最高，高達 0.505，其次為 $\log(\text{TG}/\text{HDL-C})$ 為 0.276，而在 LDL-C 則最低，只有 0.040，而利用家族相關係數估得的 HDL-C 最大遺傳率值也高達 100%。

以 HDL-C 作的混合分析(Table 3)，我們發現以 3 個成份的混合分析可以最佳解釋我們的資料，其與 4 個成份比較的 χ^2 值為 7.53，自由度為 3，P 值為 0.057，無顯著意義。若以 3 個成份與 2 個成份模型作比較，其 χ^2 值 8.02，自由度為 3，P 值為 0.046，因此我們認為 HDL-C 的分佈由 3 個成份的混合模型為最佳模型。其中平均值分別為 -0.346 0.833 4.370，而變異數為 0.516 1.043 及 0.085，而分別所作的比率為 71.5% 28.2% 及 0.3%，顯示在此族群中有一少數平均值較高，但比率很少的族群存在，其觀察值分佈與三個混合之後的觀察值分佈的圖形見圖 (Figure)。顯示期望值與觀察值分佈類似。HDL-C 分佈的情形均可以由 3 個混合分佈的模型來解釋，顯示並非只是單一的常態分佈來解釋，而其解釋最佳的模型中，也提供了進一步分析所需的參數參考，例如基因的族群頻率，平均值及變異數變等，均是相當有價值的資料。但即使由相關係數及混合分析結果顯示有明顯遺傳成份，但無法排除可能的共同環境以及生活文化型態的影響所造成的結果，因此必須利用複雜分離分析以釐清其遺傳模式，確定其為顯性或隱性，遺傳模式。

分離分析的結果(Table 4)：我們同時比較以下的模型，包括 sporadic model，多重基因模型(polygenes model)，環境 + 多重基因模型 (environmental+ Polygenes model)，主要基因模型 (pure major gene model)、主要基因+多重基因模型 (major gene + Polygenes model, i.e., mixed major gene model)，以及顯性(dominant)及隱性(recessive)基因模型，及估計全部係數的一般模型 (general model)。由於本資料以配偶對數較多($n=530$)，因此為避免參數估計不穩定，首先在多重基因模型下，先檢定模型中是 4 個家族相關係數(即 ρ_{MF} ， ρ_{MO} ， ρ_{FO} ， ρ_{SS}) 分別獨立，或是只需要 2 個係數(ρ_{MF} ， $\rho_{MO}=\rho_{FO}=\rho_{Sib}=\rho_{PO}$) 即可。而檢定結果 χ^2 值 0.94 (d.f.=2)，P 值為 0.625，因此在往後的分析均是以限制在 2 個相關係數之下進行。此減少參數數目的另一個好處，是此模型在數理論及計算上與傳統的混合模型(mixed model)是相等的(Demenis & Bonney, 1989)，而且可以用來估計遺傳率，其公式為 $h^2=2 \rho_{SS} \sigma^2 / \sigma_{\tau}^2$ ，其中 σ_{τ}^2 為總變異數， σ^2 為其模型下的變異數。利用 SAGE RegC model class D 的模型之下進行的分離分析結果如附表，若將一般模型作為基礎，以 χ^2 值計算，我們發現在 mixed environmental model, mixed co-dominant major gene model, dominant major gene

model and recessive major gene, P 值分別為 0.139, 0.734, 0.578 及 0.835，而其他的模型(sporadic, polygenes and pure major gene models) P 值分別均是 <0.001 ，顯示與一般模型比較起來，有幾個適當的模型可以解釋 HDL-C 性狀的的遺傳模型。若各種模型互相以 AIC 值作比較，發現 AIC 值在 recessive mixed major gene model 中為最低 (10431.32)，稍低於 dominant mixed major gene model (10432.75)，因此，我們可視 recessive mixed major gene model 為一較好的模型。在本研究中，我們確實可以發現有一隱性基因作用在 HDL-cholesterol 濃度的控制。其中其基因頻率為 0.280 (指控制較高 HDL-cholesterol 性狀的基因型)，三個平均值為 79.5, 50.5 及 50.5 mg/dL，而族群變異數為 127.1。配偶之間的相關係數為 0.285，而子代之間的相關係數為 0.466，估計其遺傳率為 62.7%。

本研究族群限制在健檢族群中，因此樣本以中年配偶為主要的來源，而健檢受檢者其社會經濟地位較高，無法代表一般的族群人口。但之前的研究顯示健檢受檢者仍是與動脈硬化危險因子習習相關，而且其高血壓及高脂血的盛行率與一般人口相似，因此本研究仍採取健檢患者為樣本，主要是其檢查步驟標準化，而實驗室的測量也有一致化的標準，因此可用來探索各種遺傳及環境因素，提供給臨床醫師及公共衛生學界進一步的參考。

之前的文獻回顧研究顯示血壓值及血脂質等此連續變數性狀，一方面有很多外在的境因素，如年齡、性別、體重、生活習慣等影響之外，一方面有很強的基因遺傳因素的成份。傳統的實驗室研究，常著重在一些罕見且獨特的疾病下手，例如高血脂患者則以 Familial combined hyperlipidemia 或 Familial hypercholesterolemia 等特殊疾病下手，研究其可能的基因缺陷，另外一種方式則是利用特殊的研究設計，例如雙胞胎或動物交配實驗，(作 backcross 設計)，以了解其遺傳成份的大小。本研究則是採另外一種方式去處理問題，即是利用一般族群的樣本，排除極端分布的案例，了解在一般族群中血脂及血壓的遺傳方式。一方面其研究的成果可以作為一般族群的參考，比較有公共衛生策略上的價值，一方面能解決一般族群中血脂及血壓的遺傳成份的了解，對於進一步收集族群資料及作進一步的候選基因(candidate gene markers)或是基因體全體的連鎖分析(genome wide linkage analysis)可作為重要的準備工作。

由於本研究只限制在表現型性狀的研究，並未進一步作基因型分析，但本研究指出血壓及血脂質，特別是 HDL-C 及(TG/HDL-C)比率值有明顯優先的對象，另外高血壓及血脂質的遺傳成份，也進一步聯想到胰島素抵抗性症候群可能有遺傳成份的存在，而分析只限作年齡及性別的調整，主要是避免其他因素如調整影響共同影響的基因因子的估計。

分離分析結果的討論：由於本研究的家族資料以配偶對數較多，而親子之間及手足配對較少，因此一方面模型在估計家族相關係數較能簡化成 2 個參數之外，對於其他參數似乎不太穩定，特別在傳遞參數及三個平均值參數上，不同的模型似乎相差較大。但我們仍視此分離分析的結果可以採信，其理由如下：首先，其 $-2\ln(L)$ 概似值在幾個 mixed major gene 相當接近，而

且參數估計的標準誤值均表示可以收斂，表示本資料應有足夠的樣本數目。本研究顯示有一隱性基因作用在此狀性中存在著。進一步利用候選基因方法，或是基因體整體性連鎖分析，試圖找到控制 HDL-C 性狀的基因有其必要性。而且目前已有不同的研究針對此一性狀作基因型的研究，特別是影響 lipoprotein 脂蛋白代謝過程中的酵素，如 LPL 及 CETP 等，均被報告與 HDL-C 有關的基因，另外在 Apolipoprotein 原脂蛋白的基因上，如 Apo AI/CIII/AIV 的基因簇集(gene cluster)也報告與 HDL-C 的濃度有關，因此本研究的結果可以呼應進一步的基因分析工作，同時也必須再仔細研究可能的環境因素以及可能存在的交互作用，以期能了解到控制 HDL-C 值的機制。

四、計畫成果自評

本研究是針對健檢成年人族群中所做的家族遺傳研究，特別是針對 HDL-C 的分佈情形，深入分析其家族不同配對之間的相關係數，以及其分佈的混合模型，結果顯示 HDL-C 有明顯的家族相關性及手足、親子之間的關連性，而且混合分析顯示均需超過一個常態分析的模型才能解釋，同時我們證實有一隱性基因的作用。

五、參考文獻

- Bonney, G.E. (1986) Regressive logistic models for familial disease and other binary traits. *Biometrics* 42, 611-625.
- Breslow, J.N. (1985) Human apolipoprotein molecular biology and genetic variation. *Ann. Rev. Biochem.* 54, 699-727.
- Chien, K.L., Lee, Y.T., Sung, F.C., Hsu, H.C., Su, T.C. and Lin, R.S. (1999) b Hyperinsulinemia and related atherosclerotic risk factors in the population at cardiovascular risk: community-based study. *Clin. Chem.* 45:838-846
- Demenaïs, F.M. and Bonney, G.E. (1989) Equivalence of the mixed and regressive models for genetic analysis: I. Continuous traits. *Genetic. Epidemiol.* 6, 597-617.
- Elston, R.C. and Stewart, J. (1971) A general model for the genetic analysis of pedigree data. *Hum. Hered.* 21, 523-542.
- Jarvik, G.P. (1998) Complex segregation analyses: uses and limitations. *Am. J. Hum. Genet.* 63, 942-946.
- Statistical analysis for genetic epidemiology, Release 3.1. Computer program package available from the Department of Epidemiology and Biostatistics, Rammelkamp Center for Education and Research, MetroHelath Campus, Case Western Reserve University, Cleveland. S.A.G.E. (1997)

Appendix:

Table 1. Basic demographic and atherosclerotic risk profiles in the study population of this family study

Study population (N= 1313)				
	Mean	SD	Skewness	Kurtosis
Age (years)	54.7	11.5	-0.27	0.14
Tchol (mg/dL)	194.4	36.2	0.25	1.37
TG (mg/dL)	130.4	79.3	2.26	8.05
HDL-C (mg/dL)	52.8	14.6	0.81	0.88

LDL-C (mg/dL)	116.2	31.2	0.55	0.55
Logarithm TG (mg/dL)	4.73	0.52	0.31	0.13
TG/HDL ratio	2.86	2.37	2.55	9.30

Table 2. Familial correlation coefficients in residual blood pressures and lipid profiles by FCOR programs in this family study, after adjusting with age and gender effects.

	Pairs #	HDL cholesterol	
		Equal weight to pairs	Equal weight to pedigrees
Spouse	530	0.201	0.201
Parental-Offspring	76	0.436	0.425
Sibling	52	0.505	0.518
h^2		1.165	1.194

L-C values

μ_4	Component variances				component proportions (%)				$-2\ln(l)+c$	²	P
	21	22	23	24	1	2	3	4			
	0.999				1				3725.13		
	0.562	1.316			0.775	0.225			3623.85	101.28	0.000
	0.516	1.043	0.085		0.715	0.282	0.003		3615.83	8.02	0.046
1.078	0.110	0.571	0.020	1.106	0.078	0.732	0.002	0.188	3608.30	7.53	0.057

ation analysis of residual HDL-cholesterol values: Class D regressive models

s	Environmental + Polygenes	Mixed Mendel + Polygenes	Pure Mendel	A Dominant, $\mu_{AB} = \mu_{AA}$		A Recessive,				
	0.277	0.031	0.278	0.032	0.178	0.028	0.040	0.009	0.280	0.031
	=q _A		[1]		[1]		[1]		[1]	
	=q _A		[0.5]		[0.5]		[0.5]		[0.5]	
	=q _A		[0]		[0]		[0]		[0]	
0.4	79.4	2.3	79.6	2.4	89.3	2.9	79.5	2.3	79.5	2.3
	50.6	2.8	51.7	3.2	63.3	1.8	79.5	2.3	50.5	0.5
	50.6	2.1	49.7	2.3	46.5	0.8	50.5	0.5	50.5	0.5
7.6	128.6	7.5	126.2	8.5	85.5	7.1	127.3	7.4	127.1	7.4
0.042	0.284	0.053	0.287	0.054	[0]		0.288	0.053	0.285	0.053
0.066	0.489	0.078	0.457	0.086	[0]		0.433	0.082	0.466	0.080
	7		7		5		6		6	
7	10423.37		10419.15		10444.00		10420.75		10419.32	
	5.5		1.28		26.13		2.88		1.45	
	0.139		0.734		0.000		0.578		0.835	
7	10437.37		10433.15		10454.00		10432.75		10431.32	

(Continued)

Model	General, unrestricted model	
q_A	0.268	0.035
τ_{AA}	1.000	0.000
τ_{AB}	0.645	0.127
τ_{BB}	0.000	0.000
μ_{AA}	80.0	2.7
μ_{AB}	52.5	3.9
μ_{BB}	49.2	2.5
σ^2	125.2	10.8
ρ_{MF}	0.288	0.057
$\rho_{MO}=\rho_{FO}=\rho_{SS}$	0.458	0.089
#	10	
$-2\ln(L)$	10417.87	
χ^2	Baseline	
P value		
AIC	10437.87	

Figure of HDL cholesterol distribution by commingling analysis

