

重疊病例對照和病例世代研究

于明暉

國立台灣大學 公共衛生學院 公共衛生學系與流行病學研究所

罕見慢性病的世代研究，如心臟血管疾病和癌症，需要龐大的樣本數進行長期追蹤，對於暴露因子的測量，相當耗費人力和經費，重疊病例對照研究(nested case-control study)和病例世代研究(case-cohort study)是應用在世代研究中的兩種流行病學研究設計。這兩種研究設計可以使世代研究在花費較少的狀況下進行，增加研究的效率和可行性[1-4]。

重疊病例對照研究

重疊病例對照研究結合世代研究和病例對照研究的方法，兼取兩種方法的優點。其進行過程，首先需要像世代研究一樣選擇世代，蒐集初步暴露資料(包括：病歷記錄、問卷、和生物檢體)，但這些初步所蒐集的暴露資料並不像傳統的世代研究，立即經過進一步處理或分析[1, 3, 4]。舉例來說，對飲食因子和大腸癌的重疊病例對照研究，在研究個案進入研究時即蒐集飲食問卷資料，但並不先進行這些飲食問卷資料的譯碼及鍵入電腦的工作；對於血清膽固醇和缺血性心臟病的重疊病例對照研究，在世代成員進入研究時，即收集血液檢體，但血液檢體僅經過初步前處理的過程，分離出血清後凍存在超低溫冷凍櫃中，並不先經血清膽固醇濃度的分析。重疊病例對照研究需要等到世代研究追蹤一段時間，有足夠的病例數發生後，由未發病的世代成員選取一個隨機樣本做為對照組，對照個案是在病例發生時，仍然維持無病狀態，但未來具有發病危險性的人。重疊病例對照研究僅對病例和對照組所蒐集的初步暴露

資料進一步處理，利用病例對照研究的統計方法進行分析。

重疊病例對照研究在研究個案未發病之前即蒐集暴露資料，所以暴露的測量不會受到疾病狀態的影響，譬如前述有關血清膽固醇和缺血性心臟病的研究，如果罹患缺血性心臟病之後，會使研究個案改變飲食習慣或服用降低膽固醇的藥物，利用病例對照研究法，在病例發生以後，才進行血液檢體採集和膽固醇分析，會產生血清膽固醇濃度和缺血性心臟病無顯著相關或負相關的錯誤結論。利用重疊病例對照研究設計進行研究，由於在未患病前即採集血清檢體，只是膽固醇分析是在罹病後，血清膽固醇濃度不受罹病的影響，所以在因果時序性的判定上較佳。另外；傳統的病例對照研究不易定義病例來源的族群，容易發生選擇偏差(selection bias)，重疊病例對照研究的病例組和對照組來自同一追蹤世代，所以可減少選擇偏差發生的機會。相對於傳統的世代研究，重疊病例對照研究不需要對每名世代成員初步所蒐集的暴露資料進行分析，只需分析病例組和對照組的暴露資料，所以可以節省人力和研究經費。

目前許多探討病因的流行病學研究，需要應用生物標記(biomarker)，包括：暴露、易感性、早期生物反應、及疾病標記。生物標記的測定需要較複雜精密的實驗，常是耗時和費力的工作，在人力、經費有限的狀況下，進行世代研究有所困難，所以重疊病例對照研究已經相當廣泛的應用在有關生物標記的流行病學研

Title: Nested Case-Control and Case-Cohort Study

Author: Ming-Whei Yu

School of Public Health and Graduate Institute of Epidemiology,
College of Public Health, National Taiwan University

Key Words: nested case-control study, case-cohort study

究。譬如最近已發展出酵素免疫分析法(enzyme immunoassay)測量血清中黃麴毒素 B1-白蛋白鍵結物，做為偵測個體黃麴毒素暴露的生物標記，可以用以進行黃麴毒素暴露和肝細胞癌危險性的相關性研究。肝細胞癌雖然是臺灣地區的常見癌症，但即使是屬於高危險群的男性 B 型肝炎帶原者，其每年肝細胞癌的發生率也只有 3-5% 左右，所以必須以龐大樣本數的 B 型肝炎帶原者為研究世代，一個臺灣地區有關黃麴毒素和肝細胞癌的重疊病例對照研究，世代成員包括 4841 名 30 歲以上的男性 B 型肝炎帶原者，在這個研究中，如果要對每一名世代成員皆進行黃麴毒素 B1-白蛋白鍵結物的實驗分析，將相當耗費人力和經費，為了研究的可行性，此研究將收案時的血液檢體經前處理即凍存起來，有 32 名 B 型肝炎帶原者在大約四年的追蹤期間發生肝細胞癌，對每一名肝細胞癌病例，由未罹患肝細胞癌的 B 型肝炎帶原者中，選取 2-3 名年齡(相差五歲以內)、收案來源、及血液檢體採集時間(相差三個月以內)配對的對照個案，黃麴毒素 B1-白蛋白鍵結物和麩胺基硫轉移酵素 M1、T1 基因多形性的測量，是利用 32 名病例和 73 名對照個案凍存的血清檢體和白血球 DNA，研究結果發現，在麩胺基硫轉移酵素 M1 或 T1 為無效基因型者中，血清黃麴毒素 B1-白蛋白鍵結量和肝細胞癌危險性呈顯著劑量效應關係，但對麩胺基硫轉移酵素 M1 或 T1 為非無效基因型者而言，黃麴毒素 B1-白蛋白鍵結物和肝細胞癌沒有顯著相關存在 [5]。

雖然重疊病例對照研究在因果時序性的判定，要比傳統的病例對照研究好，但對於生物標記相關的重疊病例對照研究，卻需要考慮臨床前期疾病或病理變化對生物標記的影響。舉例來說，許多病例對照研究已經發現癌症病患的血清銅濃度顯著較對照組高，發展後期的癌症病患之血清銅濃度又高於發展較前期的癌症病患，為了釐清血清銅濃度和癌症的因果關係，一個在華盛頓州的研究，利用對 5000 名世代成員凍存的血清檢體進行重疊病例對照研究，研究個案包括 133 名追蹤期間發生癌症的

病患和 241 名對照個案，研究結果發現血清銅濃度和癌症具顯著正相關，但只限於血清檢體採集時間至癌症診斷時間為四年以內的癌症病患，對於此期間為四年以上的病例和其對照個案比較，並未發現血清銅濃度和癌症危險性有關 (6)。這個研究說明了可能臨床前期癌症的發展會導致銅的吸收、代謝、和分佈改變，使得血清銅濃度偏高，所以要瞭解血清銅濃度和罹患癌症危險性的因果關係，需要追蹤一段相當長的時間，並且去除生物檢體採集以後短期內發生癌症的病例進入分析，才能避免臨床前期疾病狀態的影響。

重疊病例對照研究較傳統的世代研究在人力和經費的耗費較少，但對於生物標記的研究，需要相當大的超低溫冷凍櫃空間儲存生物檢體，建立生物檢體庫(biospecimen bank)，生物檢體的收集包括世代成員進入研究時和追蹤期間。在收集多種生物檢體(如血清，血漿，白血球，紅血球及尿液等)和同種生物檢體為了避免進行多項分析時重複解凍而需要分裝後儲存的狀況下，超低溫冷凍櫃空間的不足，常成為許多重疊病例對照研究的困難。另外，重疊病例對照研究需要考慮生物檢體中待測物質隨溫度和儲存時間改變的特性，舉例來說，血清膽固醇和三酸甘油脂可以在-20°C 的儲存狀況下，維持五年或更久的時間，不會被破壞，但血清脂蛋白(lipoproteins)卻需儲存在-70°C 的狀況下，才可以維持較久的時間。經過長期儲存的時間，等待足夠的病例數發生，常不可避免的有些生物檢體中的成份會漸漸產生改變，所以在進行生物標記的重疊病例對照研究時，對照組的選取，需要和病例組的生物檢體儲存時間相似，就成為個體匹配(individual matching)的重要條件。

重疊病例對照研究和傳統的病例對照研究同樣是以對比值(odds ratio)估計相對危險性，不能得到發生率資料，所以不能直接估計相對危險性。由於其對照個案選取的條件，必須是在病例發生時，仍然維持無病狀態者，但具有未來發病的危險性，所以可視為時間配對(time-matched)的對照個案，這樣選取對照個案



的方法，即病例對照研究中所談的密度取樣 (density sampling)。由於採用密度取樣，重疊病例對照研究可以利用對比值估計暴露組和非暴露組發生率的相對危險性 (instantaneous rate ratio)，和由世代研究直接估計相對危險性一樣，不需“罕見疾病”的假設。重疊病例對照研究的統計分析方法可以參考病例對照研究所使用的方法[7]，由於採用密度取樣，大多數學者主張直接使用配對分析的方法較為適當[4]，但實際應用上常是依照研究設計採用個體匹配與否，考慮使用配對或非配對分析，如果是在個體匹配的狀況，多變項分析可以考慮使用條件對數複迴歸模式 (conditional logistic regression model)，但如果考量配對條件不夠嚴謹或是研究設計是採非個別匹配的方式選擇對照個案，則必須使用非條件對數複迴歸模式 (unconditional regression model) 進行分析。

病例世代研究

相對於重疊病例對照研究，病例世代研究較少被使用。病例世代研究的進行，需要在世代研究一開始先選擇一個隨機樣本做為次世代 (subcohort)，次世代必須對於完整的世代具有代表性，其選擇可採隨機取樣或分層取樣，大部份先以年齡分層，再進行隨機抽樣，使得疾病好發的年齡層有足夠的樣本數，以使研究者能夠在較短的追蹤期間內獲得足夠的病例數。病例世代研究對每一世代成員進行相同的追蹤，所有的世代成員在進入研究時即接受暴露資料的蒐集，但是只有次世代成員和非次世代中追蹤期間發病的病例之暴露資料經過進一步的處理和/或分析。病例世代研究可以利用次世代監測世代研究進行的狀況，例如介入性社區實驗 (field intervention trial) 可以用次世代監測研究個案的順從性和介入對預防疾病的效果。許多世代研究可以利用次世代監測暴露資料的蒐集和分析情形，並估計暴露組和非暴露組的發生率。另外；世代研究的目的往往不只是想瞭解一種特殊健康危害的危險因子，如果想要瞭解危險因子對多種健康危害的影響時，利用

病例世代研究，只需要「單一」樣本即可進行分析，在瞭解多種健康危害的世代研究上具優越性[2-4]。

在某些情況下，病例世代研究可用以進行外在比較 (external comparison)。外在參考族群 (reference population) 常是一般具低暴露或無暴露的族群，而研究世代具有不同的暴露狀況。這個參考族群必須具有相當正確的發生率或死亡率資料，通常可應用某些已有完整且正確的疾病或死亡登記系統的族群做為外在比較的參考族群，因為在進行病例世代研究的外在比較時，可以直接利用這些族群的生命統計資料。外在比較的統計分析需要應用計算標準死亡比 (standardized mortality ratio) 的概念，可以次世代暴露組追蹤觀察的人年數估計完整的世代暴露組的觀察人年數。利用完整世代特殊年齡層總觀察人年數乘以次世代中該年齡層暴露組觀察人年數佔次世代該年齡層總觀察人年數的分率，即為完整的世代中暴露組特殊年齡層之觀察人年數的估計值，研究世代暴露組的特殊年齡層觀察人年數乘以參考族群年齡別疾病發生率，即可得出特殊年齡層期望發病人數，求出研究世代某暴露組觀察病例數和各年齡層期望病例數總和的比值，即為所估計的相對危險性指標。雖然病例世代研究方便於進行外在比較，但因為參考族群常缺乏年齡、性別以外其它重要干擾因子的資料，因此外在比較所得的結論和內在比較 (即同一追蹤世代的暴露組和非暴露組比較) 相比，仍是相當粗略的，這是在進行外在比較時重要的研究限制[4]。

病例世代研究最後進入分析的對象包括次世代和非次世代中追蹤期間發病的病例，對於生物標記的研究，雖然所有世代成員均在進入研究時，就必須接受生物檢體採集，但只有屬於次世代的研究個案之生物檢體立即進行暴露的測量，如果生物檢體中某些待測的成份容易隨儲存時間產生改變，不屬於次世代的病例，其暴露是在發病以後才分析，無法在採集生物檢體後立即測量，既然所有非病例個案因為屬於次世代，暴露的測量不受檢體儲存時間的影響，而部份病例因為不屬於次世代之故，暴露



的測量可能因檢體儲存時間較久而較易發生測量偏差，這樣將會使暴露和疾病的相對危險性估計值產生偏差，而且這個偏差是屬於系統性誤差(systematic error)。對於會增加罹病危險性的暴露，如果儲存時間會使暴露測量值降低，將會低估相對危險性，反之產生高估。相對的，重疊病例對照研究並不會有此困擾。

當然研究者也可以等待研究終止時，再將次世代和所有病例儲存的生物檢體取出分析，這樣就不會有因次世代和非次世代病例之暴露測量時間不同，帶來系統性誤差的困擾，但這樣一來，就需在研究追蹤終止時，耗費相當多的人力來配合實驗室的工作，這通常對大多數世代研究是不可行的，所以直至目前為止，病例世代研究設計使用較為頻繁的研究，仍是對暴露的測量較不受生物檢體儲存時間影響的研究。譬如在荷蘭有關硒暴露和肺癌的病例世代研究，這個研究包括 120852 名男性和女性世代成員，研究開始時，收集完初步暴露資料和生物檢體後，隨機選取 3500 名世代成員做為次世代，個體硒暴露是利用腳指甲施測，由於腳指甲硒並不會隨指甲儲存時間改變，而且可以做為累積硒暴露量指標。最後的研究結果發現，指甲硒含量和罹患肺癌危險性有顯著負相關存在，指甲硒含量最高五分位和最低五分位比較，在控制年齡、性別、抽菸及教育程度等重要危險因子後，相對危險性約 0.5 倍，而且低指甲硒含量和低 β -胡蘿蔔素及維生素 C 攝取量具交互作用[8]。值得一提的是，雖然此研究的主題是想要瞭解指甲硒和肺癌的關係，但如果此研究是以血漿或血清測 β -胡蘿蔔素和維生素 C 濃度，做為這兩種抗氧化維生素的暴露指標，既然血清(漿)中這兩種抗氧化維生素濃度會隨儲存時間改變，研究結果在探討抗氧化維生素和硒暴露的交互作用時，同樣要考慮次世代和不屬於次世代病例測量時間不同所造成的偏差。

對於病例世代研究的統計分析，較重疊病例對照研究複雜，雖然已有一些統計方法可供分析時參考[2,9]，但截至目前為止，倘未有可

用的統計分析軟體，這也是病例世代研究設計使用不如重疊病例對照研究頻繁的原因。

參考文獻

1. Kleinbaum DG, Kupper LL, Morgenstern H: *Epidemiologic Research: Principles and Quantitative Methods*. California: Wadsworth, 1982.
2. Prentice RL: A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* 1986;73:1-11.
3. Langholz B, Thomas DC: Nested case-control study and case-cohort methods of sampling from a cohort: a critical comparison. *Am J Epidemiol* 1990;131:169-76.
4. Wacholder S, Boivin JF: External comparisons with the case-cohort design. *Am J epidemiol* 1987;126:1198-209.
5. Chen CJ, Yu MW, Liaw YF, et al: Chronic hepatitis B carriers with null genotypes of glutathione S-transferase M1 and T1 polymorphisms who are exposed to aflatoxin are at increased risk of hepatocellular carcinoma. *Am J Hum Genet* 1996;59:128-34.
6. Coates RJ, Weiss NS, Daling JR, et al: Cancer risk in relation to serum copper levels. *Cancer Res* 1989;49:4353-6.
7. Breslow NE, Day NE: *Statistical Methods in Cancer Research*. vol 1. "The Analysis of Case-control Studies". Lyon: IARC Scientific Publications, 1980.
8. Van Den Brandt PA, Goldbohm RA, Van't Veer P, et al: A prospective cohort study on selenium status and the risk of lung cancer. *Cancer Res* 1993;53:4860-5.
9. Greenland S: Adjustment of risk ratios in case-base studies: hybrid epidemiologic designs *Statistics in Medicine*. 1986;5:579-84.

