

國立交通大學

統計學研究所

博士論文

邊際充分成份原因模型下辨識與估計拮抗機制

Identifying and Estimating the Antagonism with
Marginal Sufficient Component Cause Model

研究生：周孟穎

指導教授：林聖軒 博士

中華民國一百零九年五月



邊際充分成份原因模型下辨識與估計拮抗機制
Identifying and Estimating the Antagonism with
Marginal Sufficient Component Cause Model

研究生：周孟穎

Student: Meng-Ying Chou

指導教授：林聖軒

Advisor: Sheng-Hsuan Lin

國立交通大學

統計學研究所

博士論文

A Thesis

Submitted to Department of Statistics

College of Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

in

Statistics

May 2020

Hsinchu, Taiwan, Republic of China

中華民國一百零九年五月



國立交通大學

博碩士論文紙本暨電子檔著作權授權書

(提供授權人裝訂於紙本論文书名頁之次頁用)

本授權書所授權之學位論文，為本人於國立交通大學統計學研究所 _____ 組，108 學年度第 2 學期取得博士學位之論文。

論文題目：邊際充分成份原因模型下辨識與估計拮抗機制

指導教授：林聖軒

一、紙本論文授權

紙本論文依著作權法第15條第2項第3款之規定辦理，「依學位授予法撰寫之碩士、博士論文，著作人已取得學位者…推定著作人同意公開發表其著作」。

二、論文電子檔授權

本人授權將本著作以非專屬、無償授權國立交通大學、台灣聯合大學系統圖書館及國家圖書館。

論文全文上載網路公開之範圍及時間：	
中英文摘要	必需公開
本校及台灣聯合大學系統區域網路	■ 中華民國 114 年 6 月 9 日公開
校外網際網路及國家圖書館	■ 中華民國 114 年 6 月 9 日公開

說明：基於推動「資源共享、互惠合作」之理念與回饋社會及學術研究之目的，得不限地域、時間與次數，以紙本、光碟或數位化等各種方式收錄、重製與利用；於著作權法合理使用範圍內，讀者得進行線上檢索、閱覽、下載或列印。

授權人：周孟穎 (親筆簽名)

中華民國 109 年 06 月 10 日



國立交通大學研究所博士班
論文口試委員會審定書

本校統計學研究所 周孟穎 君所提論文

邊際充分成份原因模型下辨識與估計拮抗機制

Identifying and Estimating the Antagonism with

Marginal Sufficient Component Cause Model (題目)

合於博士資格水準、業經本委員會評審認可。

口試委員：黃冠華 林端祥
黃介標 楊怡暉
林聖軒

指導教授：林聖軒 教授

研究所所長：沈聖合 教授

中華民國 109 年 5 月 25 日



邊際充分成份原因模型下辨識與估計拮抗機制

研究生：周孟穎

指導教授：林聖軒 博士

國立交通大學統計學研究所

摘要

充分成份原因是因果推論領域用來探討發生結果的可能原因的方法。在流行病學上為了減少疾病的發生，需要考慮造成疾病的可能原因。而充分成份原因適合用來描述這些原因與風險因子之間的作用機制。許多研究文獻由充分成份原因的架構下，提出方法來檢定兩個因子的交互作用。並且也有學者把充分成份原因和反事實結果結合，在經驗法則之下來估計各種機制的大小。然而現有的估計方法建立在因子對結果有沒有預防效果或是單調遞增的條件的假設，在此假設之下，估計的機制只可以一併估計協同作用與促效作用，或是由對二元變數重新編碼來估計拮抗作用。同樣的在檢定交互作用的方法，協同作用與拮抗作用無法同時被檢定。當拮抗作用與協同作用同時存在時，需要其它方法來一起估計與檢定。因此單調遞增假設下，無法由現有的方法中把拮抗機制從協同與促效機制辨識出來。

本研究經由去除單調遞增假設來延伸邊際充分成份原因模型，考慮因子對結果的影響有預防效果的可能，推論可以產生的邊際機制。本研究方法可以對邊際拮抗機制、邊際促效機制、邊際協同機制三種交互作用存在的情況

下，辨識出拮抗機制，也同時估計與檢定三種機制。在考慮以上機制交互作用可能的七種組合情況，模擬分析的結果顯示誤差相當小，覆蓋率信賴區間的覆蓋率接近 95%。此結果顯示如果資料的結構和 mSCC 模型的設定相同，本研究方法可以用來作為估計與檢定以上三種機制交互作用。於資料分析部分我們採用台灣世代研究的肝癌資料。考慮兩個因子的拮抗、促效與偕同機制存在的情況下使用 mSCC 模型來分析，結果顯示拮抗交互作用的檢定結果顯著，也就是說在 HCV 抑制 HBV 的機制然後導致肝癌發生可以經由 mSCC 模型來說明。

中文關鍵字：邊際充分成份原因模型、充分成份原因模型、反事實結果、因子有預防效果、單調遞增假設。

英文關鍵字：Marginal sufficient component cause (mSCC), sufficient component cause (SCC), counterfactual outcome, no preventive effect assumption, monotonicity assumption.

Identifying and Estimating the Antagonism with Marginal Sufficient Component Cause Model

Students: Meng-Ying Chou

Advisor : Sheng-Hsuan Lin

Institute of Statistics
National Chiao-Tung University

Abstract

Sufficient component cause (SCC) is the principle concept in causal inference research field. SCC is applied when one investigates the causes of a result.. In epidemiology, to reduce the prevalence of a disease, it is necessary to consider the possible causes. There had been many literatures which provided methods for testing the effect of two variables interaction. Some researchers integrate the SCC and counterfactual methods to measure and estimate the mechanistic interaction effect.

However these methods were based on two assumptions. One assumption in SCC scenario was risk factors had no preventive effect on the outcome. The other in counterfactual outcome scenario was the counterfactual outcome had monotonicity with respect to risk factors. Under the assumptions, there were only synergistic mechanism and agonistic mechanism. To estimate the antagonistic mechanism, one of the binary factor had to be recorded. The synergistic and agonistic mechanism could not be tested along with the antagonistic mechanism at the same time when the factor

was recorded. It required another way to identify and test the agonistic and synergistic mechanism at the same time when the both mechanism was considered co-existed.

In this research, we extended the marginal sufficient component cause (mSCC) model by considering two risk factors had preventive effects on the response variable. We inferred the possible mechanism with preventive effect setting in mSCC model. With this setting, the marginal antagonism, agonism and antagonism could be identified with surrogate variables. They could be tested and estimated at the same time. We perform simulation to validate our method. The biases approximated to zero and the coverage rates were closed to 95% in each of the seven scenarios from the three mechanistic interactions combinations. And we applied the mSCC model to evaluate the antagonism in the Taiwanese cohort study of HCC data. The result suggested that there was significant antagonism among HCV and HBV that led to HCC when considering the three mechanisms.

Keywords: Marginal sufficient component cause (mSCC), sufficient component cause (SCC), counterfactual outcome, no preventive effect assumption, monotonicity assumption.

致謝

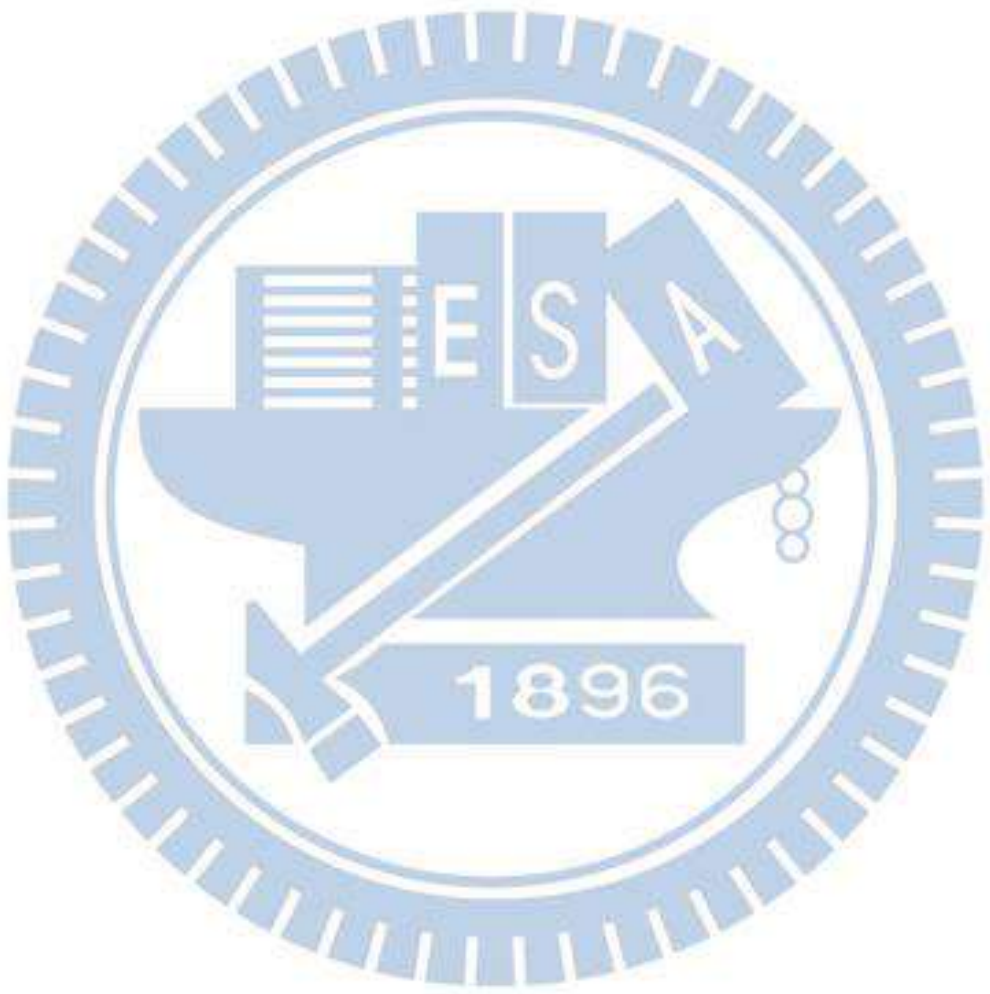
能夠完成博士學位，需要感謝很多人。論文能夠完成，我要感謝林聖軒老師的耐心，總是為我仔細說明因果推論與流行病學這兩的學門的方法。也在研究論文撰寫、投影片的呈現、學術研究方法與生涯等給予我相當多寶貴的建議。我要感謝戴安順博士協助幫忙校正論文初稿與演練投影片的報告。我也要感謝黃冠華老師的耐心，於我詢問學術上的問題時總是能給予我中肯的建議。我感到榮幸也要感謝黃彥棕老師、楊懷壹老師、林瑞祥老師、黃冠華老師、林聖軒老師擔任我的學位考試委員，給與我的研究的建議並且提供研發方法未來可延伸的方向。

在就讀博士期間，我也必須要感謝盧鴻興老師與黃冠華老師讓我有機會參與所上的交通資訊預測計畫，讓我從參與當中瞭解老師們如何建立預測模型的過程，也讓我了解同學們如何去實作模型方法。我要感謝盧鴻興老師與陳素雲老師幫我撰寫博士班入學推薦函，讓我符合參加入學考試的條件，也讓我從工作回到學術研究。我也要感謝杜憶萍老師在我大學畢業時提供了研究助理工作機會，讓我有機會走入統計的學術領域，繼續學習機率統計方法。

我要感謝林聖軒老師、黃冠華老師、王維菁在我就讀博士班期間，提供我研究津貼的支持，讓我可以全力專心的研究統計方法與理論。也要感謝統計所提供的各種助教工作，除了提供而外的收入也讓我從中獲得統計方法與程式設計的教學經驗。我也要感謝統計的行政人員們的幫忙，解決我遇到的各種申請文件疑問。同時也要感謝所上的老師們，在我遇到您們時總是願意回答我詢問的問題。我也要感謝在博士班期間的同學、學長姐與學弟妹們，與你們談話的過程中，我知道了最新流行的、許多有趣的事物或資料分析方法與工具。

最後我要感謝我的父母在我從高中到博士班的這段期間，讓我自由探索數學方法和讓我選擇進入研究所就讀，也感謝我的家人們與親戚們的對我的支持。也要感謝我女朋友在這段日子的支持。有大家的支持我才有辦法完成博士班學業。

孟穎 2020/06



目錄

中文摘要.....	i
英文摘要.....	iii
致謝.....	v
目錄.....	vii
表目錄：.....	ix
圖目錄：.....	xi
1. 介紹.....	1
1.1. 統計推論方法與因果推論方法.....	1
1.2. 充分成份原因與機制檢定.....	2
1.3. 機制估計.....	3
1.4. 邊際充分原因(mSCC).....	4
1.5. 研究問題與目標.....	5
1.6. 資料分析對象.....	6
1.7. 研究結果.....	7
2. 文獻方法回顧與本研究貢獻.....	8
2.1. 反事實結果與充分成份原因.....	9
2.1.1. 反事實結果.....	10
2.1.2. 充分成份原因.....	11
2.1.3. 拮抗機制交互作用的例子.....	12
2.1.4. 充分成份原因與反事實結果.....	13
2.2. 檢定交互作用.....	15
2.2.1. 風險尺度下.....	15
2.2.2. 存活函數尺度下.....	17
2.3. 機制大小的估計.....	19
2.4. 邊際充分成份原因.....	22
2.5. 研究目標.....	25
3. 本研究的方法.....	26
3.1. 反應模式對應到充分成份原因.....	27
3.2. 考慮 B 型與 C 型肝炎病毒的拮抗機制.....	31

3.2.1.	充分成份原因框架下的拮抗機制.....	31
3.2.2.	反應模式與邊際充分成份原因.....	32
3.2.3.	mSCC 的背景條件與 SCC 背景條件的對應	36
3.3.	邊際充分成份原因模型的拮抗機制.....	38
3.3.1.	對邊際充分成份原因的機率模型.....	38
3.3.2.	代理變數的定義和性質.....	41
3.4.	辨識拮抗作用的機制.....	43
3.5.	估計.....	46
4.	數值分析.....	49
4.1.	模擬分析.....	49
4.1.1.	考慮拮抗、促效、協同三種機制.....	50
4.1.2.	考慮交互作用只有拮抗與協同兩種機制的存在.....	51
4.1.3.	假設只有協同與促效互作用機制.....	53
4.1.4.	假設只有拮抗與促效.....	55
4.1.5.	假設只有拮抗.....	57
4.1.6.	假設只有促效.....	59
4.1.7.	假設只有協同.....	60
4.2.	資料分析.....	62
4.2.1.	只考慮協同與促效兩種機制.....	64
4.2.2.	拮抗、協同與促效三種機制.....	65
5.	討論.....	68
	附錄：資料模擬的生成細節.....	71
	參考文獻.....	76

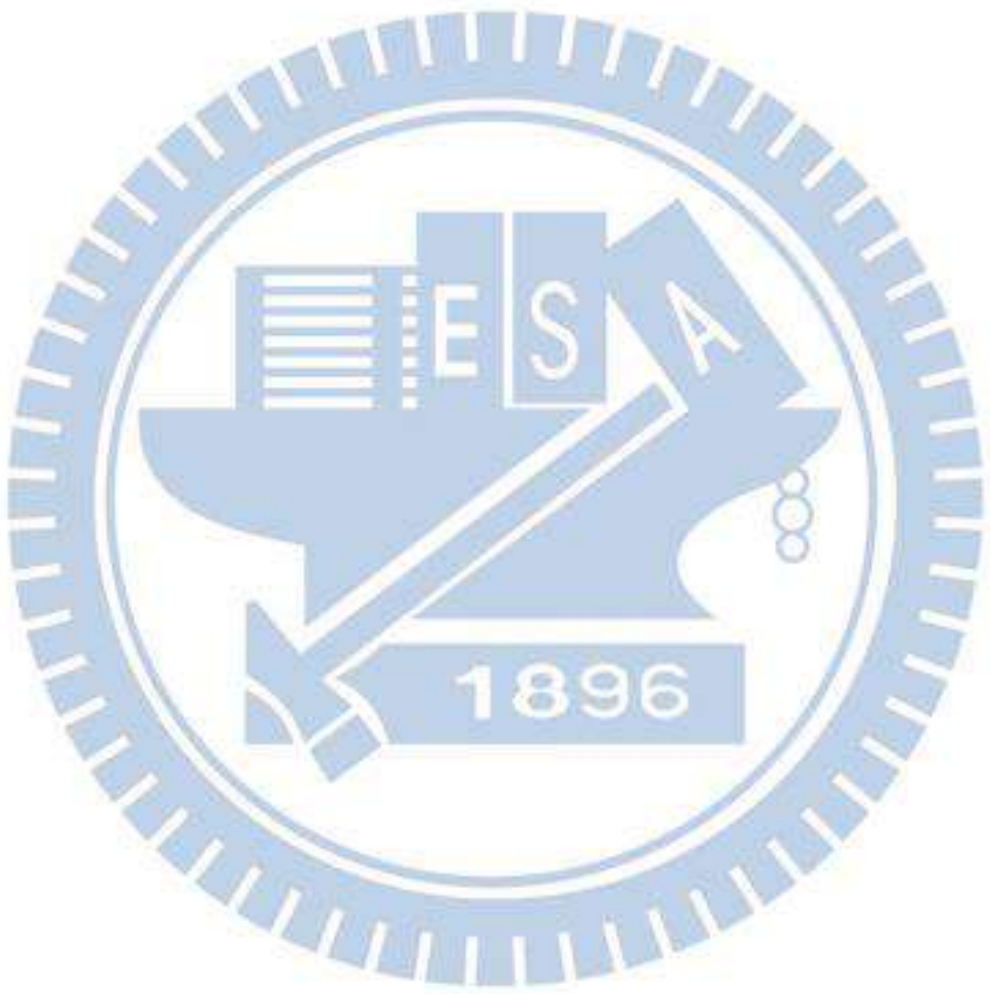
表目錄：

表格 1：在充分成份原因模型下，假設兩個因子對結果的影響沒有預防性效果，背景條件構成的向量與反應模式的對應。.....	24
表格 2：在充分成份原因模型，考慮拮抗、協同兩種機制交互作用存在的狀況下，背景成分組成的向量與反應模式的對應關係。.....	33
表格 3：由充分成份原因的背景條件，考慮背景條件的存在與不存在，以集合的非補集與補集表示(欄位左 2)，然後對應到有關的風險因子(欄位右 2)。最後歸納出反應模式(欄位左 1)與邊際成分原因(欄位右 1)。.....	35
表格 4：在邊際成分原因模型下考慮拮抗、協同兩種機制交互作用，對應到在邊際成份原因模型下的拮抗、協同、促效三種邊際機制交互作用。由左而右分別是反應模式、邊際效果的因子、MSCC 的背景成分、機制的理論值、機制的估計值。.....	48
表格 5：拮抗、促效、協同三種機制交互作用同時存在的模擬結果。表格欄位由左而右分別是理論機率值、比例、估計機率值、偏差值、估計標準差、95%信賴區間上下界、模擬覆蓋率。.....	51
表格 6：拮抗與協同二種機制交互作用同時存在的模擬結果。表格欄位由左而右分別是理論機率值、比例、估計機率值、偏差值、估計標準差、95%信賴區間上下界、模擬覆蓋率。.....	53
表格 7：協同與促效二種機制交互作用同時存在的模擬結果。表格欄位由左而右分別是理論機率值、比例、估計機率值、偏差值、估計標準差、95%信賴區間上下界、模擬覆蓋率。.....	54
表格 8：拮抗與促效二種機制交互作用同時存在的模擬結果。表格欄位由左而右分別是理論機率值、比例、估計機率值、偏差值、估計標準差、95%信賴區間上下界、模擬覆蓋率。.....	56
表格 9：只有拮抗機制交互作用的模擬結果。表格欄位由左而右分別是理論機率值、比例、估計機率值、偏差值、估計標準差、95%信賴區間上下界、模擬覆蓋率。.....	58

表格 10：只有促效機制交互作用的模擬結果。表格欄位由左而右分別是理論機 率值、比例、估計機率值、偏差值、估計標準差、95%信賴區間上下界、模擬 覆蓋率。.....	60
表格 11：只有協同機制交互作用的模擬結果。表格欄位由左而右分別是理論機 率值、比例、估計機率值、偏差值、估計標準差、95%信賴區間上下界、模擬 覆蓋率。.....	61
表格 12：考慮 HCV 與 HBV 導致 HCC 經由促效、協同兩種機制存在的情況下，由 邊際成份原因模型所估計的各機制大小。欄位由左而右分別是機率、比例、估 計誤差、95%信賴區間上下界、P 值。.....	64
表格 13：考慮 HCV 與 HBV 導致 HCC 經由拮抗、促效、協同三種機制存在的場合 下，由邊際成份原因模型所估計的各機制大小。欄位由左而右分別是機率、比 例、估計誤差、95%信賴區間上下界、P 值。.....	66
表格 14：比較有拮抗機制(欄位左)與沒有拮抗機制(欄位右)的邊際機制的背景成 份理論值大小。可以發現在沒有拮抗機制的場合下(欄位右)，拮抗機制的期望 值會在因子 X2 的效果與協同機制效果這兩者分別出現。.....	68

圖目錄：

圖 1：兩個因子，背景條件因子，邊際結果，干擾因子等四者之間的因果非有向循環圖。.....	44
圖 2：兩個因子，背景條件因子，代理變數，邊際結果，干擾因子等五者之間的因果非有向循環圖。.....	45
圖 3：拮抗、促效、協同三種機制交互作用同時存在情況下，模擬的偏差值 95%信賴區間。.....	51
圖 4：拮抗與協同二種機制交互作用同時存在情況下，模擬的偏差值 95%信賴區間。.....	53
圖 5：協同與促效二種機制交互作用同時存在情況下，模擬的偏差值 95%信賴區間。.....	55
圖 6：拮抗與促效二種機制交互作用同時存在情況下，模擬的偏差值 95%信賴區間。.....	57
圖 7：只有拮抗機制交互作用情況下，模擬的偏差值 95%信賴區間。.....	59
圖 8：只有促效機制交互作用情況下，模擬的偏差值 95%信賴區間。.....	60
圖 9：只有協同機制同時存在情況下，模擬的偏差值 95%信賴區間。.....	62
圖 10：考慮 HCV 與 HBV 導致 HCC 經由促效、協同兩種機制作用存在情況下，由邊際成份原因模型估計的 95%信賴區間。.....	65
圖 11：考慮 HCV 與 HBV 導致 HCC 經由拮抗、促效、協同三種機制作用存在情況下，由邊際成份原因模型估計的 95%信賴區間。.....	67



1. 介紹

1.1. 統計推論方法與因果推論方法

兩個變數的關係可以經由統計模型來建立，而兩變數之間關聯的大小可以統計模型背後的機率來描述。關聯的大小通常經由統計相關係數來描述，然而變數的相關(correlation)不代表兩個變數之間存在著因果效應(causal effect)。即使在應變數(response variable)與自變數(independent variable)之間建立了高度相關的機率統計預測模型(probabilistic and statistical prediction model)，變數之間的因果關係仍須經由因果推論(causal inference)方法來解釋。因果推論的非有向循環圖(direct acyclic graphs)適合用來解釋變數之間的因果關係，也是用來判斷干擾因子的方法。在去除了所有可能的干擾因子的影響之後，才可以說明變數之間的統計相關的確存在因果效應。非有向循環圖只是因果推論的一種方法，其它方法包括了反事實結果(counterfactual factual)、中介因子分析(mediation analysis)、路徑效用分析(path analysis)、交互作用分析(causal interaction analysis)。在這些分析方法當中，可再依據研究的目標分成兩類，一類是對因果效用大小的估計，另一類是結果的背後機制探討。

學者與資料分析師往往有興趣於觀察到的現象與它發生的原因。分析瞭解結果背後的原因與確立因果關係，有助於在未來相似的情況下去推論各種結果的可能性。反事實結果與充分成份原因 (sufficient component cause, SCC)是在因果推

論裡面兩個重要的方法。前者讓學者與資料分析師量化並估計結果的大小。後者被用來討論與呈現結果背後的各種可能機制。探討機制的 SCC 模型在研究文獻上，相對於反事實結果模型的比例只佔非常小的一部分[1]。雖然學者 VanderWeele 指出探討結果背後機制的方法只在相關理論學門的課堂上講述，即使目前研究目標大多著眼在反事實結果方法上，他認為探討結果的背後機制仍不可以忽略[1]。

1.2. 充分成份原因與機制檢定

Rothman [2]的充分成份原因概念可以幫助我們了解結果發生的背後可能機制。在充分成份原因的概念之下，結果的發生由背後的多種機制是否完成而決定。每個機制形成到完成之前，必須先有一系列的必需的條件成份(necessary component)。當機制的這些必須條件都具備後，該機制就會驅動結果發生。為了更明瞭辨別兩個風險因子之間的作用而導致結果發生的各種交互作用機制，過去的學者們發展出一系列的方法，來辨識、檢定、估計這些結果發生的可能機制。

交互作用可以初步分類為兩種，一個是兩個風險因子對結果的影響可能有超可加性質(superadditivity)，另一個是次可加性質(subadditivity)。前者為兩個因子的共同作用效果大於個別因子效果的總和，後者為兩個因子的共同效果小於個別因子效果的總和。VanderWeele, et al. [3]對機制交互作用(interaction)做定義與分類，他們依據上述兩種性質，在充分成份原因概念裡把交互作用分類為協同機制的交互作用(synergistic interaction)與拮抗機制的交互作用(antagonistic interaction)。

以兩個變數代表風險因子，充分成份原因的協同作用的呈現，必須是當結果發生時背後機制有兩個風險因子存在且作用。充分成份原因的拮抗作用的呈現，必須是當結果發生時背後機制有兩個風險因子存在，其中一個因子必須存在並且作用，另一個因子必須存在但是不作用。檢定協同作用是否存在，經由計算結果變數的風險比率(risk ratio)或勝算比率(odds ratio)，再對二元變數的兩個風險因子建立線性飽和迴歸模型(linear saturated regression model)，看兩個變數項的係數是否顯著，來判斷協同機制是否存在。如果要判斷拮抗機制是否存在，可以對兩個風險因子的其中一個二元變數重新編碼，再對兩個變數項的係數檢定，由它來判斷拮抗機制的存在。學者 Lee [4]提出 PRISM 統計量，他把資料的時間加入考慮，建立存活函數與兩個風險因子的模型，來檢定正向協同機制作用和負向協同機制作用。PRISM 統計量建立在存活函數的尺度下來對交互作用檢定。當交互作用存在時，相對在不考慮時間的風險比率、勝算比率的尺度，PRISM 更能反應交互作用的存在。

1.3. 機制估計

相對於機制的檢定，部分研究文獻在充分成份原因概念之下去估計結果背後的各個機制。學者 Hoffmann, et al. [5]過去曾嘗試藉由可歸因分數(population attributable fraction)，簡寫 PAF，來量化結果發生背後的充分成份原因機制。他提出了可歸於充分之因類別的疾病比例 (proportion of disease due to a class of

sufficient causes, PDC), 藉由集合的運算, 透過 PAF 來呈現兩個因子對結果的充分成份原因的四個機制, 並把機制的呈現與估計推展到兩個因子以上的情況。然而該估計方法並未達到目的估計原因論的歸因分率(etiology attributable fraction), 學者 VanderWeele Tyler [6]說明學者 Hoffman 的估計方法只可以估計額外歸因分率(excess attributable fraction), 而且只可以用在因子對結果的影響為單調遞增的情況下。在系統化估計 SCC 機制的研究部分, 學者 Liao, et al. [7]在 SCC 概念下兩個風險子作用的結果的機制量化, 他們提出因果圓派加權(causal pie weight, CPW) 的概念, 由統計的多變數線性回歸模型(multiple linear regression model)來估計充分成份原因的機制。此方法由多變數線性回歸模型所預測的 odds ratio 來估計 PAF, 並且在兩個風險因子的情況下把 PAF 與 CPW 用矩陣呈現並做連結。學者 Lee [8]提出潛在完成(complete potential, CP)的概念, 藉由在多變數線性回歸模型估計的比率(rate ratio)來估計 CP, 在定義 rate profile (RP), 藉由勝算比(odds ratio)近似比率, 然後對因果圓派加權 CPW。此研究對 CPW 的估計方法做了一般化的建構。

1.4. 邊際充分原因(mSCC)

以上的方法對於機制的檢定與估計必須分開進行, 而建立估計機制的過程缺少因果推論的辨識步驟。學者 Lin, et al. [9]對結果背後的機制提出了可以同時估計與檢定的方法。為了討論 HCV 與 HCV 在 HCC 的促效機制交互作用, 他們提

出了邊際充分成份原因模型(marginal sufficient component causes)，簡稱 mSCC。此方法包括了兩個概念，把量化因果效應的方法和結果發生背後原因的方法結合。也就是把反事實結果的方法，和充分成份原因的方法對應。把各機制組合對應到反事實結果定義的反應模式，再由統計方法估計各個反應模式對應的邊際機制效果，同時推論機制的存在與否。在 mSCC 模型下邊際機制，延伸了一種在 SCC 模型無法表達的促效機制(agonistic mechanism)。為了估計促效機制，學者 Lin, et al. [9]提出了估計四種機制交互作用的方法。依照線性模型辨識機制的需要在模型中加入代理變數，來估計與量化協同和促效兩種機制。此方法同時估計兩種機制作用是過去研究沒有的。此模型的最大特色是可以同時估計與檢定風險因子的邊際效果。之後學者 Lin, et al. [10]考慮結果變數在不同時間標記下的紀錄，把 Lin, et al. [9]的模型延伸到存活分析的風險函數上。

1.5. 研究問題與目標

以上的檢定或估計的方法雖然可辨別交互作用是否存在，但是這些方法都是在充分組成原因概念下，建立在兩個風險因子對結果的影響都沒有預防效果的情況。雖然檢定拮抗作用可以由二元變數的風險因子重新編碼，經由檢定協同作用的方法來達成，然而檢定的方法無法同時估計該機制的大小。雖然 CPW 的方法嘗試估計 SCC 背後交互作用機制的量，但是此方法仍限制在兩個風險因子對結果的影響，是單調遞增的假設之下，也就是固定一個風險因子的值，增加另一個

風險因子的值，會使增加後的結果值，比風險因子增加前的結果值，比較之下前者大於等於後者。邊際充分成份原因模型可以估計邊際協同作用，還有在充分成份原因模型無法描述的促效作用。然而目前的邊際充分成份原因模型的方法，也是建立在結果變數對風險因子是單調遞增的情況下。拮抗機制的估計必須沒有在此假設之下才會在充分成份原因模型裡面被呈現。或是在單調假設下的邊際充分成份原因模型，把兩個二元風險因子的其中一個重新編碼，由 1 轉換成 0、0 轉換成 1，如此一來才可以估計拮抗機制。然而重新對變數編碼後，可能導致原本想估計的機制無法一併被同時量化。本研究將會釋放單調遞增假設，討論在釋放假設之後的邊際充分成份原因模型可以估計的三種機制交互作用，並且引入代理變數來估計拮抗機制，讓模型一併估計和檢定協同與促效等共三種機制。

1.6. 資料分析對象

由釋放單調遞增假設的 mSCC 模型，我們用此方法來分析台灣肝癌資料庫在 1991、1992 這兩年蒐集的世代研究資料。Hepatocellular Carcinogenesis (HCC) 是在成年人當中常見的一種肝癌，它的發生通常會經由肝硬化。在所有可能造成 HCC 的風險因子當中，B 型和 C 型肝炎病毒是引發 HCC 高風險的因子[11-15]。當個體只感染 B 肝病毒，或是只感染 C 肝病毒，或是先後感染兩種病毒，會使肝細胞有很高的可能性發展成癌細胞。然而這兩個風險因子對 HCC 的交互作用，除了協同作用和促效作用[9]，仍有可能存在拮抗作用。從學者 Huang, et al. [16]

的研究顯示，把 B 肝病毒 DNA 數量當作應變數，C 肝病毒 RNA 的數量當作自變數，在考慮調整其它變數後的結果，呈現負相關。由於在世代研究上，感染 B 肝病毒的個體為在出生前和出生後的幾年內，而感染 C 肝病毒的時間點為個體成長一段時間後。加上 C 肝病毒的 RNA 製造的蛋白質會抑制 B 肝病毒的 DNA 數量，因此這兩種病毒對肝癌除了可能經由促效機制，也有可能經由拮抗機制的交互作用導致肝癌發生。因此本研究將釋放原本在邊際充分原因模型的單調假設條件來延伸估計方法，並且經由模擬驗證與資料分析，來實現辨識、估計與檢定以上三種機制。

1.7. 研究結果

方法模擬部分我們由估計模型對應的機率分配模擬生成資料，再由本研究提出的方法估計三種機制。資料生成的部分，採用白努力分配依各自設定的參數產生兩個風險因子、代理變數、干擾因子等數值。邊際成份原因的各機制背景成份，由多項分配模擬產生數值。然後我們比較各機制的理論值與模型的估計值，模擬結果顯示兩者的數值相當接近。估計方法對於理論值的覆蓋率，也與統計上大樣本性質的 95% 覆蓋率相當接近。這兩點驗證了本研究提出的方法。

資料分析部分，我們分析 B 型與 C 型肝炎在邊際成份原因模型下，六種邊際機制導致肝癌的大小。這六種機制分別如以下描述：由 B 肝與 C 肝以外的原因導致肝癌發生的機制、只由 B 肝病毒使肝細胞直接突變成癌細胞的機制、只由

C 肝病毒產生發炎環境而導致肝細胞突變成癌細胞的機制、在 C 肝病毒產生的發炎環境下與 B 肝病毒使肝細胞突變的作用，加速癌細胞產生的協同機制、C 肝病毒 RNA 抑制 B 肝病毒 DNA 的拮抗機制、B 肝病毒與 C 肝病毒競爭同一肝細胞進而使其突變成癌細胞的促效機制。從分析結果可以推論，造成肝癌的機制，除了透過這兩個風險因子以外的機制效果是顯著的，兩個各別風險因子導致結果的機制，拮抗機制也是顯著的。而且在疾病的個體當中，拮抗機制的比例與協同機制的比例相當接近。代表 B 型與 C 型肝炎透過拮抗機制導致肝癌發生的路徑是不可忽略的。

2. 文獻方法回顧與本研究貢獻

在這一章會先我們回顧目前在因果推論領域的兩個方法類別，一個是研究結果機制的充分成份原因模型，另一個是反事實結果概念。然後介紹在結果變數在單調遞增假設下的邊際充分成份原因模型。說明在此假設之下，如何把充分成份原因模型的機制與反事實結果結合。接著在下一章介紹本研究，在不考慮單調假設的邊際充分成份原因模型的機制，在同樣概念的對應下，會拓展多少個邊際機制。

為了說明過去文獻，與本研究使用的估計或檢定方法，在此將會使用下列符號於式子當中。結果變數的部分， Y 代表觀察到的結果， $Y(x_1, x_2)$ 代表在因果推論的反事實結果變數， Y_V 代表在充分成份原因的結果變數， Y 代表在邊際充分成

份原因的結果變數。 X_1 與 X_2 代表會影上述四種變數的兩個風險因子。 x_1, x_2 分別代表在因果推論的反事實結果變數 $Y(x_1, x_2)$ ，它的兩個風險因子 X_1 與 X_2 介入(intervene)的值。 p_{ij} 代表給定風險因子 $X_1 = i$ 與 $X_2 = j$ 的條件下， $Y = 1$ 的機率，它等於條件期望值 $E[Y|X_1 = i, X_2 = j]$ 。

在充分成份原因模型的符號， E_j 代表一個事件的完成。 V_j 代表在充分成份原因的模型下第 j 個背景成份因。背景成份原因是在一系列的必要條件，它們使結果發生的成份，扣除一個風險因子 X_1 或 X_2 ，或是扣除兩者，剩餘的其它條件。為了一致呈現過去文獻的方法，以 V_0 背景條件對應和風險因子無關的機制($E_0 = V_0$)、 V_1 背景條件對應和風險因子 X_1 有關的機制($E_1 = V_1X_1$)、 V_2 背景條件對應和風險因子 X_2 有關的機制($E_2 = V_2X_2$)、 V_3 背景條件對應和風險因子 X_1 、 X_2 有關的協同機制($E_3 = V_3X_1X_2$)。

於邊際成份原因模型的符號部分， S_j 代表邊際充分成份原因的一個事件的完成，以 U_j 代表在邊際成份原因模型下的邊際背景成份因，其中 U_0 背景條件對應和風險因子無關的邊際機制($S_0 = U_0$)、 U_1 背景條件對應和風險因子 X_1 有關的邊際機制($S_1 = U_1X_1$)、 U_2 背景條件對應和風險因子 X_2 有關的邊際機制($S_2 = U_2X_2$)、 U_3 背景條件對應和風險因子 X_1 、 X_2 有關的邊際協同機制($S_3 = U_3X_1X_2$)。 U_4 背景條件對應和風險因子 X_1 、 X_2 有關的邊際促效機制($S_4 = U_4(X_1 + X_2 - X_1X_2)$)。

2.1. 反事實結果與充分成份原因

2.1.1. 反事實結果

在因果推論的研究領域，學者 Splawa-Neyman, et al. [17]提出反事實結果的概念。近年的研究，學者 Pearl, et al. [18]描述了觀察資料、介入原因、考慮反事實結果這三項是在因果推論裡重要的概念。評估變數間的因果關係，可以先經由觀察資料，然後思考並假設變數之間可能的關聯，再決定因子與結果後，可以介入因子變數，最後考慮可能發生的反事實結果並量化它。理論上我們想由介入的方式改變原先觀察到的因子，然後比較因子在介入前與介入後的反事實結果，來推論事先猜想的原因與結果之間是否有因果關係。然而反事實結果無法直接觀察到，這是因為觀察到某一個體的結果 $Y_k(x_1, x_2)$ ，人們無法經由時間回溯改變風險因子 X_1, X_2 的值，使它們由 $X_1 = x_1, X_2 = x_2$ 改變成 $X_1 = x'_1, X_2 = x'_2$ ，所以結果 $Y_k(x'_1, x'_2)$ 將無法得知。因此實際上在評估因果效應會藉由估計反事實結果的期望值 $E[Y(x_1, x_2)]$ ，來討論變數與因子之間的因果關係。在可交換(exchangeability)條件與一致(consistency)條件的假設下，可以藉由觀察到變數的條件期望值 $E[Y|X_1 = x_1, X_2 = x_2]$ 估計反事實結果的期望值。由此步驟我們可以比較因子在不同介入值得結果，來判斷因子與變數之間的因果關係是否存在，也可以估計它的平均大小。

隨機試驗的資料是經由估計反事實結果的平均值，來評估因果關係的最好方法[19, 20]。經由隨機分配讓各干擾因子分層的機率分配一致，達到去除干擾因子對因子與結果之間因果關係的影響。在隨機分配下蒐集到的資料，給定的因

子的值 $X_1 = x_1, X_2 = x_2$ 所對應的變數 Y 的機率分配，因此而接近反事實結果 $Y(x_1, x_2)$ 的機率分配，因此由資料得到的統計相關可以解釋成因果關係。

2.1.2. 充分成份原因

反事實結果的概念可以讓我們瞭解兩個以上變數之間的因果關係。然而要討論造成結果的背後的各種可能原因或是可能的機制，必須經由充分成份原因的概念。研究文獻中最早對充分成份原因這個概念做系統性的說明與完整的介紹，來自學者 Rothman [2]的研究。他的充分成份原因概念用來呈現結果變數發生的可能原因。一個觀察到的結果，由一系列的事件來表示。這一系列的事件是對於結果的充分非必要條件，只要有一個事件發生就會使結果 Y_V 發生。其中每一個事件的發生，必須由一個以上的必要成份組合而成，事件要在所有成份都存在的前提下，事件才會發生。如果組成事件的所有成份當中有一個成份不存在，事件便不會發生。藉由充分成份原因的概念讓學者能夠明瞭與系統化呈現結果變數 Y_V 發生的可能原因與成份。假設事件 E_i 的發生必須由兩個成份作用才會發生，那麼把機制的其中兩個成份用 X_1 、 X_2 代表呈現且有作用，用 X_1^C 、 X_2^C 代表呈現但是不會作用。當結果 Y_V 的發生對應到事件 E_1 發生時，事件 E_1 的兩個成份 X_1 、 X_2 都必須存在且有作用，則 E_1 機制稱為協同機制交互作用(synergistic interaction)。如果結果 Y_V 的發生對應到事件 E_2 發生時，事件 E_2 的兩個成份 X_1 、 X_2 ，其中一個成份存在且有作用，以 X_1 表示；另一個成份必須存在但是不作用，以 X_2^C 表示，那麼 E_2 機制稱為

拮抗機制交互作用(antagonistic interaction)。相同的概念，如果 E_3 機制對應的必須成份為 X_1^c 、 X_2 ，則 E_3 也稱為拮抗機制交互作用。

2.1.3. 拮抗機制交互作用的例子

學者 Mutlu, et al. [21] 示範三個在臨床研究上拮抗機制交互作用的實際範例。

在這裡用變數 X_1, X_2 分別代表事件的兩個成份，變數 Y_V 代表結果。

第一個範例是注意力缺凡過動症 X_1 、藥用安非他命 X_2 的兩個成份影響過動症 Y_V 結果發生的機制。一個體如果有注意力缺凡過動症的($X_1 = 1, X_2 = 0$) 會有過動症 ($Y_V = 1$)。沒有過動症的個體接受藥物安非他命治療 ($X_1 = 0, X_2 = 1$)會產生過動現象 ($Y_V = 1$)。但是在有注意力缺凡過動症的個體會使用藥物安非他命治療 ($X_1 = 1, X_2 = 1$)，則個體過動的情況會被治療 ($Y_V = 0$)。從第一個範例可以觀察到不管是 X_1 、 X_2^c 或是 X_1^c 、 X_2 都可以得到 Y_V 的結果，因此可以得知注意力缺凡過動症 X_1 、藥用安非他命 X_2 兩者會經由拮抗機制交互作用導致過動症 Y_V 發生。

第二個範例是低氧濃度環境 X_1 、高濃度一氧化碳環境 X_2 這兩個成份對個體的紅血球細胞的氧飽和濃度 Y_V 的影響。個體只暴露在低氧濃度環境 之下($X_1 = 1, X_2 = 0$)，假設環境中只有氧氣沒有一氧化碳，那麼個體的紅血球細胞的含氧濃度充足 ($Y_V = 1$)。個體只暴露高濃度的一氧化碳環境下($X_1 = 0, X_2 = 1$)，個體的紅血球細胞會吸收一氧化碳，使得紅血球細胞的含氧濃度不足($Y_V = 0$)。如果個體同時暴露在低氧濃度與高濃度的一氧化碳環境下($X_1 = 1, X_2 = 1$)，細胞和氧分子結合的位置會有很大的傾向被一氧化碳取代，造成個體的紅血球細胞的含氧

濃度不足 ($Y_V = 0$)。因此從第二個範例可以觀查到結果 Y_V 會由 X_1 、 X_2^c 的作用而發生，因此只有低濃度氧氣的環境 X_1 和只有高濃度一氧化碳環境 X_2 ，對紅血球細胞的氧氣飽和度是否充足 Y_V 有拮抗機機制交互作用。

第三個範例是高血壓 X_1 、血管硬化 X_2 對認知力下降 Y_V 的影響。如果個體只有高血壓($X_1 = 1, X_2 = 0$)，比正常情況下較有可能會導致認知力下降 ($Y_V = 1$)。如果個體只有血管硬化($X_1 = 0, X_2 = 1$)，也會有比較高的風險發生認知力下降 ($Y_V = 1$)。然而有高血壓的個體($X_1 = 1$)，會有較高可能使得血管硬化($X_2 = 1$)，使個體的認知力下降的風險降低($Y_V = 0$)。從這個例子也可以觀察到 X_1 、 X_2^c 或是 X_1^c 、 X_2 都可以對應到 $Y_V = 1$ 的結果。因此可以說明高血壓 X_1 與血管硬化 X_2 對認知力下降的風險 Y_V 存在拮抗機機制交互作用。以上三個情況不但給予了實際例子，也說明了拮抗機機制可以經由對變數的定義從資料觀察到。

2.1.4. 充分成份原因與反事實結果

充分成份原因提供了方法來呈現結果背後的可能機制，為了要檢定和估計事件的各種機制或交互作用，必須經由量化的資料與反事實結果概念，由統計推論和因果推論來估計。學者 Flanders [22]把充分成份原因模型，結合反事實結果方法一併呈現。並在考慮學者 Greenland, et al. [23]所提出的兩個二元風險因子之下對應的反事實結果，考慮四種因子的組合，討論四個反事實結果元素的向量。此向量之為反應模式(response pattern)。他提出把結果對應到反應模式來估計風險

因子在充分原因模型下可能的事件機制大小。同一年稍後的研究文獻，學者 VanderWeele, et al. [24]對 Flander 的方法提出需要修正的地方，並且示範把反事實結果 $Y(x_1, x_2)$ ，連結充分原因模型的結果 Y_V ，把充分原因模型的結果透過可能事件與二元的析取或(disjunctive or)運算子呈現[25]。假設事件對應的成份 X_1 、 X_2 分別用 1 代表發生，0 代表不發生。析取或運算子 \vee 的定義如以下所描述， X_1 、 X_2 上運算的結果則是 $X_1 \vee X_2 = 1$ 當 $X_1 = 1$ 或是 $X_2 = 1$ 的時候，其它情況則為 0。假設結果變數的發生只有兩個事件發生，則可以析取或運算子表示為 $Y_V = E_1 \vee E_2$ 。如果事件 E_1 與 E_2 分別由 $V_1A_1A_2$ 與 $V_2A_1^cA_2$ 所組成，那麼結果 Y 可以進一步表示為 $Y_V = E_1 \vee E_2 = V_1A_1A_2 \vee V_2A_1^cA_2$ 。

考慮學者 Suzuki, et al. [26]提出的超額分數(excess fraction)、可歸因分數(attributable fraction)、原因分數(etiology fraction)，該文獻指出考慮反事實結果的單調遞增假設和 SCC 模型下的單調遞增假設應該做區分，在區分兩者假設的不同下，會有上述三種估計因子對結果的效果。根據該文獻的定義，在充分成份原因的結果 Y_V 如果有單調遞增，則對應的風險因子對結果的影響為沒有預防效果(no preventive action)。在此假設下。因子在兩個因子之下結果變數 Y_V ，可以經由析取或運算子與可能的九種機制： $E_0 = V_0$ 、 $E_1 = V_1X_1$ 、 $E_2 = V_2X_2$ 、 $E_3 = V_3X_1^c$ 、 $E_4 = V_4X_2^c$ 、 $E_5 = V_5X_1X_2$ 、 $E_6 = V_6X_1^cX_2$ 、 $E_7 = V_7X_1X_2^c$ 、 $E_8 = V_8X_1^cX_2^c$ 來表示。在兩個因子對結果的影響為有預防效果的假設下，這九種機制會縮減為 $E_0 = V_0$ 、 $E_1 = V_1X_1$ 、 $E_2 = V_2X_2$ 、 $E_5 = V_5X_1X_2$ 四種機制。此時的結果可以表示成 $Y_V = E_0 \vee$

$E_1 \vee E_2 \vee E_5$ ，代表 Y_V 的結果，至少由其中一個以上的事件發生。它可以再被表示成 $Y_V = V_0 \vee V_1X_1 \vee V_2X_2 \vee V_5X_1X_2$ 。如果考慮每一個事件發生的時間，學者 Lee [27] 提出了無多餘假設(no redundancy assumption)，在極短的時間下只會有一個事件使結果 Y_V 發生。因此在這個條件之下 $Y_V = E_0 + E_1 + E_2 + E_5$ ，也就是 $Y_V = V_0 + V_1X_1 + V_2X_2 + V_5X_1X_2$ 。這式子代表 Y_V 的結果只會由其中一個事件發生而出現。這兩種理論都對在充分成份原因的結果做了與反事實結果的連結。因此在充分成份原因的結果 Y_V ，可以加上因果推論的介入概念，用反事實結果 $Y_V(x_1, x_2)$ 來呈現。

2.2. 檢定交互作用

2.2.1. 風險尺度下

拮抗機制的結果為兩個因子經由交互作用產生結果中的其中一種機制，因此在概念上檢定拮抗機制是否存在，可以用 VanderWeele, et al. [3], [28, 29]提出的方法，以判斷協同機制交互作用的方法，經由對變數重新編碼後，來檢定拮抗機制交互作用。考慮兩個因子的反事實結果 $Y(x_1, x_2)$ ，並且控制干擾因子 C ，在因果推論的可交換條件： $Y(x_1, x_2) \perp X_1|C$ 、 $Y(x_1, x_2) \perp X_2|C$ ，和一致假設之下： $E[Y(x_1, x_2)|C] = E[Y|X_1 = x_1, X_2 = x_2, C]$ ，反事實結果的期望值 $E[Y(x_1, x_2)] = E_C[E[Y|X_1 = x_1, X_2 = x_2, C]]$ ，觀察變數的期望值。把反事實結果經由期望值使其量化，可以達到檢定交互作用於方法上所需要的估計值。為了說明 VanderWeele 的檢定交互作用方法，在這裡比照過去文獻，假設 X_1 、 X_2 對 Y 的效果不受任何干

擾因子影響。因此可以先定義 $E[Y|X_1 = x_1, X_2 = x_2] = p_{x_1x_2}$ ，分別給予符號 $p_{11} = E[Y|X_1 = 1, X_2 = 1]$ 、 $p_{10} = E[Y|X_1 = 1, X_2 = 0]$ 、 $p_{01} = E[Y|X_1 = 0, X_2 = 1]$ 、 $p_{00} = E[Y|X_1 = 0, X_2 = 0]$ 。

個別因子的效用可以定義為： X_1 的效用 $p_{10} - p_{00}$ ， X_2 的效用 $p_{01} - p_{00}$ ，兩個因子的共同效用 $p_{11} - p_{00}$ 。如果有交互作用存在，那麼兩個因子的效用不會等於個別因子的效用加總，也就是 $p_{11} - p_{00} \neq (p_{10} - p_{00}) + (p_{01} - p_{00})$ 。如果反事實結果對兩個因子都是單調遞增，那麼當兩個因子的效用扣掉個別因子效用後如果不等於零，當 $p_{11} - p_{00} - [(p_{10} - p_{00}) + (p_{01} - p_{00})] = p_{11} - p_{10} - p_{01} + p_{00} \neq 0$ ，可以判斷有交互作用存在。如果反事實結果只對一個因子是單調遞增，則 $p_{11} - p_{10} - p_{01} > 0$ 的時候可以判斷有交互作用存在；如果沒有單調遞增，則當 $p_{11} - p_{10} - p_{01} - p_{00} > 0$ 的時候，可以判斷兩個因子對結果有交互作用。

由於在定義上拮抗機制所導致的結果，必須是一個因子存在並且有作用，另一個因子必須存在，但是不作用，也就是說它是機制作用當中的必須成份。所以拮抗機制在歸類上，它是負向的交互作用。假設兩個因子對結果的拮抗機制作用條件，分別是因子 X_1 必須存在，但是不作用，因子 X_2 存在並且有作用。那麼當 $p_{01} - p_{00} - p_{11} - p_{10} > 0$ 的時候，可以判斷有拮抗機制的交互作用存在。如果反事實結果對重編碼(把 0 換成 1、1 換成 0)之後的二元 X_1 因子有單調遞增(也就是反事實結果對 X_1 因子有單調遞減)，或是對另一因子 X_2 因子有單調遞增，在任一種或是兩種都成立情況之下，當 $p_{01} - p_{00} - p_{11} > 0$ 的時候，可以判斷有拮抗機制的

交互作用。如果反事實結果對重編碼的 X_1 因子和 X_2 因子都有單調遞增，那麼可以用比較強的條件，當 $p_{01} - p_{00} - p_{11} + p_{10} > 0$ 的時候，可以判斷有拮抗機制的交互作用存在。在以上情況的 $p_{x_1x_2}$ 可以由兩個自變數的線性模型估計，其中結果和自變數的關係會經由廣義線性模型的鏈結函數做連結。

在對等鏈結函數下(identity link)，假設 $E[Y|X_1 = x_1, X_2 = x_2] = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_{12}x_1x_2$ ，比照帶入因子後的係數與 $p_{x_1x_2}$ ，可以得到 $p_{00} = \beta_0$ 、 $p_{10} = \beta_1$ 、 $p_{01} = \beta_2$ 、 $p_{11} = \beta_{12}$ 。因此檢定拮抗機制的交互作用的三個式子，可以分別寫成 $\beta_2 - \beta_0 - \beta_{12} - \beta_1 > 0$ 、 $\beta_2 - \beta_0 - \beta_{12} > 0$ 、 $\beta_2 - \beta_0 - \beta_{12} + \beta_1 > 0$ 。

在對羅吉斯鏈結函數下(logit link function)，假設 $\text{logit}(E[Y|X_1 = x_1, X_2 = x_2]) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_{12}x_1x_2$ ，同樣的概念可以得到 $p_{00} = \frac{\exp(\beta_0)}{1+\exp(\beta_0)}$ 、 $p_{10} = \frac{\exp(\beta_0+\beta_1)}{1+\exp(\beta_0+\beta_1)}$ 、 $p_{01} = \frac{\exp(\beta_0+\beta_2)}{1+\exp(\beta_0+\beta_2)}$ 、 $p_{11} = \frac{\exp(\beta_0+\beta_1+\beta_2+\beta_{12})}{1+\exp(\beta_0+\beta_1+\beta_2+\beta_{12})}$ 。因此可以依照反事實結果對因子的單調遞增與否情形，用以下三個的式子來判斷拮抗作用是否存在。

$$\frac{\exp(\beta_0+\beta_2)}{1+\exp(\beta_0+\beta_2)} - \frac{\exp(\beta_0)}{1+\exp(\beta_0)} - \frac{\exp(\beta_0+\beta_1+\beta_2+\beta_{12})}{1+\exp(\beta_0+\beta_1+\beta_2+\beta_{12})} - \frac{\exp(\beta_0+\beta_1)}{1+\exp(\beta_0+\beta_1)} > 0$$

$$\frac{\exp(\beta_0+\beta_2)}{1+\exp(\beta_0+\beta_2)} - \frac{\exp(\beta_0)}{1+\exp(\beta_0)} - \frac{\exp(\beta_0+\beta_1+\beta_2+\beta_{12})}{1+\exp(\beta_0+\beta_1+\beta_2+\beta_{12})} > 0$$

$$\frac{\exp(\beta_0+\beta_2)}{1+\exp(\beta_0+\beta_2)} - \frac{\exp(\beta_0)}{1+\exp(\beta_0)} - \frac{\exp(\beta_0+\beta_1+\beta_2+\beta_{12})}{1+\exp(\beta_0+\beta_1+\beta_2+\beta_{12})} + \frac{\exp(\beta_0+\beta_1)}{1+\exp(\beta_0+\beta_1)} > 0$$

2.2.2. 存活函數尺度下

比較 VanderWeele 的以上檢定方法，它們是在風險的尺度之下(risk scale)，不考慮時間變數，來檢定交互作用，Lee [4]提出了 peril ratio index synergy multiplicity (PRISM)檢定統計量。*PRISM*的建構需要一系列的步驟。給定兩個因子的值，它對給定的風險函數在時間上去做平均，先轉換成與存活函數(survival function)相同的單位。然後平均後的量稱作危害因子 $Peril^{x_1x_2}$ 。接著為了只考慮因子的效果， $Peril^{x_1x_2}$ 會經由除以 $Peril^{x_1=0,x_2=0}$ 調整，調整後的值叫做危害比率(peril ratio)， $PR^{x_1x_2} = \frac{Peril^{x_1x_2}}{Peril^{00}}$ 。最後比較兩個因子同時作用和個別作用的效果， $PRISM = \frac{PR^{11}}{PR^{10}PR^{01}}$ 。*PRISM*可以視為在存活函數的尺度下檢定交互作用，它的優勢在於式子形式，和過去使用風險比率(risk ratio)的多重性(multiplicity)檢定交互作用式子形式上相同，但是*PRISM*的背後加入了時間的元素。使用*PRISM*檢定交互作用的時候，必須先確定假設檢定。檢定交互作用是否存在，虛無假設和對立假設分別是 $H_0: PRISM = 1$ ， $H_1: PRISM \neq 1$ 。確定交互作用存在後，再檢定是正向或負向的交互作用。檢定正向交互作用的虛無假設和對立假設分別是 $H_0: PRISM \leq 1$ ， $H_1: PRISM > 1$ 。檢定負向交互作用的虛無假設和對立假設分別是 $H_0: PRISM \geq 1$ ， $H_1: PRISM < 1$ 。*PRISM*除了可以判斷交互作用是否顯著，它可以判斷在 SCC 模型之下的協同機制、拮抗機制是否顯著。判斷協同機制的假設檢定是， $H_0: PRISM \leq Peril^{00}$ ， $H_1: PRISM > Peril^{00}$ 。判斷 $X_1 = 0, X_2 = 1$ 拮抗機制的假設檢定是， $H_0: PRISM \geq \frac{1}{Peril^{10}}$ ， $H_1: PRISM < \frac{1}{Peril^{10}}$ 。判斷 $X_1 = 1, X_2 = 0$ 拮抗機制的假設檢定是， $H_0: PRISM \geq \frac{1}{Peril^{01}}$ ， $H_1: PRISM < \frac{1}{Peril^{01}}$ 。判斷 $X_1 =$

0 $X_2 = 0$ 的因子重編碼之下的協同機制的假設檢定是， $H_0: PRISM \leq Peril^{11}$ ，

$H_1: PRISM > Peril^{11}$ 。

2.3. 機制大小的估計

以上的檢定方法聚焦在判斷交互作用是否存在，尚未量化結果背後的機制。

在一個因子與結果變數的情況下，學者 Suzuki, et al. [30]分別從反事實結果、充

分成份原因、考慮順序的充分成份原因這三個角度，說明超額分數、歸因分數、

原因分數這三個估計量。假設變數 E 代表充分成份原因的事件，變數 Y 代表觀察

到的結果， $Y(1)$ 和 $Y(0)$ 分別代表因子在 $X = 1$ 和 $X = 0$ 的反事實結果。超額分數等

於 $\frac{Pr(Y(1)=1)-Pr(Y(0)=0)}{Pr(Y(1)=1)}$ ，它呈現如果從 $X = 1$ 改變成 $X = 0$ ，在反事實結果之下，因

子改變後，群體的超額風險相對於暴露風險的比值。歸因分數等於

$\frac{Pr(Y=1)-Pr(Y(0)=0)}{Pr(Y=1)}$ ，它呈現了相對於觀察到群體在 $Y = 1$ 結果的風險之下，觀察到

$Y = 1$ 結果的風險，和在 $X = 0$ 的反事實結果的風險差異。原因分數加入了時間與

充分成份原因機制的考慮。假設一個因子 X 影響結果發生 $Y = 1$ 的充分成份原因

機制有三種，分別是 $E_0 = V_0, E_1 = XV_1, E_2 = X^cV_2$ ，它們自對應可能發生的時間分

別為 d_ϕ, d_X, d_{X^c} 。再假設 t 是觀察到 $Y = 1$ 的時間，學者 Suzuki, et al. [30]定義總和

原因分值(total etiology fraction)等於 $\frac{Pr[\{d_X \leq t\}, \{d_X < d_\phi\} | X = 1]Pr[X=1]}{Pr[Y=1]}$ ，它代表

在實際觀察到的案例，有多少的比例實際上是由因子 X 所引起。原因分數的概念

準確地描述結果發生的背後機制的測量大小，然而它無法由資料計算，只能以歸

因分數作為它的下界值。另外在因果推論的可交換假設和一致假設成立下，超額分數與歸因分數會相等，計算上會以 $\frac{\Pr(Y = 1|X = 1) - \Pr(Y = 1|X = 0)}{\Pr(Y = 1|X = 1)}$ 來估計它們。而這個式子過去文獻上的人群歸因分數 population attributable fraction(PAF)。

學者 Hoffmann, et al. [5] 討論兩個因子作用下的結果，並且量化充分成份原因的各機制。他把量化的充分成份原因機制的命名為 proportion of disease due to a class of sufficient cause(PDC)，意思是疾病的發生，有多少比例是出自於某些充分成份原因機制的類別。這些充分成份原因機制類別，被歸類對應的各風險因子所造成的結果，對應可能機制的集合。PDC 的估計的由人群歸因分數 population attributable fraction(PAF)轉換計算。PAF 的計算方法為比較因子存在結果和因子不存在的結果的差異，看這個差異在所有結果中所佔的比例。假設 Y 表示結果， X 表示因子，那麼 $PAF(X) = \frac{\Pr(Y = 1|X = 1) - \Pr(Y = 1|X = 0)}{\Pr(Y = 1|X = 1)}$ 。

因此因子 X_1 的人群歸因分數 PAF 會是所有 PDC 中，包含任何有 X_1 存在的機制。所以在兩個因子的情況下，它會包含只有 X_1 因子的機制，也包含同時有兩個因子 X_1, X_2 的機制，用式子表示可以寫成 $PAF(X_1) = PDC(X_1) + PDC(X_1, X_2)$ 。用相同的概念，因子 X_2 的人群歸因分數 PAF 的式子可以表示成 $PAF(X_2) = PDC(X_2) + PDC(X_1, X_2)$ 。兩個因子的人群歸因分數可以表示成 $PAF(X_1, X_2) = PDC(X_1) + PDC(X_2) + PDC(X_1, X_2)$ 。因此各因子下的 PDC 可以由 PAF 來表示。
 $PDC(X_1) = PAF(X_1, X_2) - PAF(X_2)$ 、 $PDC(X_2) = PAF(X_1, X_2) - PAF(X_1)$ 、
 $PDC(X_1, X_2) = PAF(X_1, X_2) - PAF(X_1) - PAF(X_2)$ 。學者 Hoffman 提出了估計因

子機制的方法，並且推廣到兩個因子以上的情況。然而由於 PAF 的計算定義為

$$PAF(X) = \frac{\Pr(Y = 1|X = 1) - \Pr(Y = 1|X = 0)}{\Pr(Y = 1|X = 1)}$$
，因此範圍可能落在 $(0, 1)$ 區間之外。

為了使估計落在理論值的範圍，Hoffman 提出用 $\Pr(X = 1|Y = 1)$ 乘以一個比例數

字來估計 $PDC(X)$ 。而這個比例數字由 $PAF(X)$ 來決定。所以由 X 因子引發的機制，

它的估計會被表示成

$$PDC(X) = \Pr(X = 1|Y = 1) \frac{\Pr(Y = 1|X = 1) - \Pr(Y = 1|X = 0)}{\Pr(Y = 1|X = 1)}$$
。

雖然學者 Hoffman 提出對機制的估計可以由 PAF 獲得，但是估計方法背後缺少對於機制的辨別。學者 Liao, et al. [7] 提出了加權因果圓派的 causal pie

weighting(CPW)估計充分成份原因的機制。兩個二元因子的組合可以對應四個

PAF，PAF 與 CPW 的關係分別是 $PAF_{none} = CPW_{00}$ 、 $PAF(X_1) = CPW_{00} + CPW_{10}$ 、

$PAF(X_2) = CPW_{00} + CPW_{01}$ 、 $PAF(X_1, X_2) = CPW_{00} + CPW_{10} + CPW_{01} + CPW_{11}$ 。

與 Hoffman 不同的地方在 PAF 的估計。Liao 提出的 PAF 是由資料計算的勝率比

odds ratio、模型估計的勝率比 odds ratio 所呈現。而模型的勝率比(odds ratio)經

由對勝率(odds)來表現，估計勝率的式子為 $odds = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_{12} x_1 x_2$ 。

Liao, et al. [7] 提出的方法可以把 PAF 與 CPW 以聯立方程式呈現，然後再透過矩

陣關係表示，然而關於如何把估計的勝率(odds)轉換為目標要估計的勝率比(odds

ratio)的細節未於文獻中明確說明。

為了補足這一部分，學者 Lee [8] 提出用 completion potential (CP) 的概念經由

建立勝率 odds 與因子的關係來估計 CPW。相對於學者 Hoffman 和 Liao 在估計機制時的 PAF 是建立在發生率(risk)的尺度，學者 Lee 的 CP 建立在風險(rate)尺度。經由 CP 呈現的風險比例(rate ratio)可以在勝率(odds)和因子之間建立線性模型，模型當中各項因子的係數則對應到充分成份原因機制的 CP。最後 CPW 再由對應到的 CP，除以所有可能機制的 CP。學者 Lee 的方法說明了估計方法中在勝率(odds)與 CPW 的連結，然後再由 CPW 回推到 PAF。

2.4. 邊際充分成份原因

以上許多學者們分別提出 PDC、CPW、CP 這三種概念來估計因子作用在結過得背後機制，然而這些估計的過程當中仍缺乏在模型裡對於機制辨識 (identification)的步驟。學者 Lin, et al. [9]提出邊際充分成份原因模型辨識機制，並且估計促效機制作用與協同機制作用。邊際充分成份原因模型把討論機制的充分成份原因模型連結到反事實結果，藉由估計每一個反應模式(response pattern)所對應的風險因子的邊際效果大小，來量化分類過後的一組充分成份原因，它們所形成的邊際機制。而這些分類的每一組充分成份原因，稱之為邊際充分成份原因。邊際充分成份原因模型的建立有三個步驟。先考慮所有機制形成時它們必須的成份的背景條件，然後把背景條件的組合以二元變數表示並且以向量呈現。接著把向量呈現的二元變數背景條件的轉換成反應模式，把每個背景條件的組合，對應到它的反應模式。最後依據每個反應模式對應的背景條件組合來做為邊際

機制作用。

反應模式以向量的方式，描述在所有因子的可能組合之下對應的反事實結果。學者 Greenland, et al. [23] 為了分類兩個因子對結果的影響、因子之間的交互作用，提出了反應模式。考慮兩個二元因子 X_1 、 X_2 ，反應模式定義為四種因子組合的反事實結果向量 $(Y(X_1 = 1, X_2 = 1), Y(X_1 = 1, X_2 = 0), Y(X_1 = 0, X_2 = 1), Y(X_1 = 0, X_2 = 0))$ 。為了方便表示，簡化記做 $(Y(1, 1), Y(1, 0), Y(0, 1), Y(0, 0))$ 。當反事實結果 $Y(x_1, x_2)$ 是一個二元變數的時候，反應模式有 $2^4 = 16$ 可能的組合。

在因子對結果的影響沒有預防效果的假設之下，結果可以由四個充分成份原因的機制與勝率運算子表示， $Y_V = V_0 \vee X_1 V_1 \vee X_2 V_2 \vee X_1 X_2 V_3$ 。考慮充分成份原因的反事實結果， $Y(x_1, x_2)$ 會由二元因子 X_1 、 X_2 ，還有二元變數的背景條件 V_0, V_1, V_2, V_3 所決定。由二元變數構成的背景成份向量 (V_0, V_1, V_2, V_3) 有 16 種可能。把這十六個可能與因子帶入 $Y_V = V_0 \vee X_1 V_1 \vee X_2 V_2 \vee X_1 X_2 V_3$ ，然後將 Y_V 、 X_1 、 X_2 與反事實結果 $Y(x_1, x_2)$ 對應，可以得到這 16 個向量對應的反應模式。它們分別是 $(0, 0, 0, 0)$ 、 $(1, 1, 0, 0)$ 、 $(1, 0, 1, 0)$ 、 $(1, 0, 0, 0)$ 、 $(1, 1, 1, 0)$ 、 $(1, 1, 1, 1)$ 。因此可以知道在因子對結果沒有預防作用的假設前提下，16 種反應模式會簡化為 6 種。若去觀察反應模式與對應的背景成份向量，反應模式 $(0, 0, 0, 0)$ 對應到一個向量，這個反應模式代表結果不發生。反應模式 $(1, 1, 0, 0)$ 對應到兩個向量，而這兩個向量可以說明這個反應模式代表因子 X_1 的邊際結果。反應模式 $(1, 0, 1, 0)$ 對應到兩個向量，這兩個向量可以說明這個反應模式代表因子 X_2 的邊際結果。反應模式

(1, 0, 0, 0)對應到一個向量，這個向量可以說明這個反應模式代表因子 X_1 、 X_2 的協同機制。反應模式(1, 1, 1, 0)對應到兩個向量，這兩個向量的組成都有 V_1 、 V_2 的背景成份，可以說明這個反應模式代表因子 X_1 、 X_2 的促效機制。反應模式(1, 1, 1, 1)對應到八個向量，這個向量可以說明這個反應模式下的反事實結果總是會發生。詳細的背景向量對應反應模式可以參考表格 1。

表格 1：在充分成份原因模型下，假設兩個因子對結果的影響沒有預防性效果，背景條件構成的向量與反應模式的對應。

Response pattern ($Y(1, 1), Y(1, 0), Y(0, 1), Y(0, 0)$)	Background condition vector (V_0, V_1, V_2, V_3)
(1, 1, 1, 1)	(1, 1, 1, 1), (1, 1, 0, 0), (1, 0, 1, 1), (1, 0, 0, 1), (1, 1, 1, 0), (1, 1, 0, 0), (1, 0, 1, 0), (1, 0, 0, 0)
(1, 1, 0, 0)	(0, 1, 0, 1), (0, 1, 0, 0)
(1, 0, 1, 0)	(0, 0, 1, 1), (0, 0, 1, 0)
(1, 0, 0, 0)	(0, 0, 0, 1)
(1, 1, 1, 0)	(0, 1, 1, 1), (0, 1, 1, 0)
(0, 0, 0, 0)	(0, 0, 0, 0)

考慮反應模式與各自對應的若干個背景成份向量，可以把反映模式的結果用對應的邊際背景條件與因子呈現， $Y = U_0 + U_1X_1 + U_2X_2 + U_3X_1X_2 + U_4(X_1 + X_2 - X_1X_2)$ 。為了與充分成份原因模型區分，在這裡 Y 代表邊際成份原因模型的結果， U_j 代表邊際機制的背景成份。 U_0 對應反應模式(1, 1, 1, 1)、 U_1 對應反應模式(1, 1, 0, 0)、 U_2 對應反應模式(1, 0, 1, 0)、 U_3 對應反應模式(1, 0, 0, 0)、 U_4 對應反應

模式(1, 1, 1, 0)。在估計邊際機制大小的方法，假設 U_3 、 U_4 之中只有一個機制會出現，則其它的機制可以用兩個自變數的線性飽和模型(linear saturated model)的係數估計。如果 U_3 、 U_4 對應的機制都存在，必須在模型中增加自由度才有辦法估計。學者 Lin, et al. [9]提出代理變數(surrogate variable)的概念解決自由度不足的問題。代理變數必須滿足兩個條件，一個是它必須是充分成份原因機制的必要成份，另一個是他可以是結果發生的風險因子、或者也可以不是結果發生的風險因子。代理變數是在性質上會滿足生物交互作用機制中的接收子(receptor)的生物標籤(biomarker)。在 Lin, et al. [9]的這個研究中代理變數為 U_4 背景條件當中的一個成份，因此上述的式子會被表示成 $Y = U_0 + U_1X_1 + U_2X_2 + U_3X_1X_2 + U_4Z_1(X_1 + X_2 - X_1X_2)$ ，其中 $U_4 = U_{40}Z_1$ 。在機制估計的部分，將會分別考慮代理變數不是風險因子的線性模型 $E[Y(x_1, x_2, z_1)] = p_0 + p_1x_1 + p_2x_2 + p_3x_1x_2 + p_{40}z_1(x_1 + x_2 - x_1x_2)$ ，或是考慮代理變數是風險因子的線性模型 $E[Y(x_1, x_2, z_1)] = p_0 + p_1x_1 + p_2x_2 + p_3x_1x_2 + p_{40}z_1(x_1 + x_2 - x_1x_2) + p_{z_1}z_1 + p_{z_2}z_1x_1 + p_{z_3}z_1x_2$ 。學者 Lin, et al. [10]把反事實結果發生的時間加入考慮，把邊際充分成份原因模型用來估計風險函數，經由轉換成存活函數，克服了直接用線性模型估計二元變數的反事實結果的機率，可能發生模型應變數超出範圍的問題。

2.5. 研究目標

以上的方法在研究交互作用機制有檢定與估計兩類。在檢定的方法上，可以個別分開檢定協同作用與拮抗作用，在估計方可以測量一個因子的機制或是兩個因子的協同作用機制。然而目前的估計方法只能估計因子對結果沒有預防效果的機制，然而根據文獻的例子，與關於檢定拮抗機制或者負向交互作用的方法，有必要再發展方法來辨識與估計拮抗作用。因此本研究將延伸邊際充分原因模型，採用它對於機制可辨識的優特點，把在 SCC 模型下因子的無預防效果假設(單調遞增假設)去除，建立可以同時估計拮抗作用、協同作用、促效作用的模型。然後我們將以肝癌資料庫來示範本研究方法的估計三種機制的估計與檢定結果。

3. 本研究的方法

在第二章的部分介紹了反應模式，它考慮了所有可能的風險因子的組合情況，列表所有可能的反事實結果。在不考慮任何關於風險因子與反事實結果的假設，反事實結果組成的四個變數的向量 $(Y(1,1), Y(1,0), Y(0,1), Y(0,0))$ 的反應模式，總共會有十六種。如果假設充分成份原因的兩個因子 X_1 、 X_2 的機制對結果沒有預防效果，充分成份原因的機制可以簡化為 4 種，在經過背景成份四個元素的向量 (V_0, V_1, V_2, V_3) 和兩個因子 X_1 、 X_2 的組合轉換成反應模式之後，反應模式可以簡化為六種。如果兩個因子 X_1 、 X_2 的機制對結果有存在預防效果，充分成份原因的機制會有九種，而這九種機制的背景成份元素的向量，會構成 $2^9 = 512$ 種的背景成份組合，然後再與因子 X_1 、 X_2 轉換成 16 個反應模式。以下的內容將逐步說明如

何把512種的背景成份組合的向量歸類到 16 個反應模式。並且考慮肝癌資料的風險因子對結果的機制，從 9 種簡化到 5 種。並且說明在邊際充分成份原因模型的反應模式對應到的拮抗作用機制，以及模型的機制辨識與估計。

本研究會沿用第二章的符號，與下列的額外符號於式子當中。 V_j 代表在 SCC 模型下的背景成份因，其中 V_3 背景條件對應和風險因子 X_1^C 有關機制、 V_4 背景條件對應和風險因子 X_2^C 有關機制、 V_5 背景條件對應和風險因子 X_1 、 X_2 有關的協同機制、 V_6 背景條件對應和風險因子 X_1 、 X_2^C 有關的拮抗機制、 V_7 背景條件對應和風險因子 X_1^C 、 X_2 有關的拮抗機制、 V_8 背景條件對應和風險因子 X_1^C 、 X_2^C 有關的機制。 U_j 代表在 mSCC 模型下的背景成份因，其中 U_0 背景條件對應和風險因子無關的邊際機制、 U_1 背景條件對應和風險因子 X_1 有關的邊際機制、 U_2 背景條件對應和風險因子 X_2 有關的邊際機制、 U_3 背景條件對應和風險因子 X_1 、 X_2 有關的邊際協同機制、 U_4 背景條件對應和風險因子 X_1 、 X_2 有關的邊際拮抗機制、 U_5 背景條件對應和風險因子 X_1 、 X_2 有關的邊際促效機制、 Z_1 為辨識拮抗機制的代理變數、 Z_2 為變數促效機制的代理變數。

3.1. 反應模式對應到充分成份原因

在充分成份原因的概念裡，結果變數 Y_V 的發生可能是由一個或是兩個以上的機制來表示。考慮二元結果變數 Y_V 和二元風險因子變數 X_1, X_2 ，在風險因子可能有預防效果的情況下， Y_V 的結果可能由充分成份原因的九種機制構成。以符號

代表分別是 $E_0 = V_0$ 、 $E_1 = V_1X_1$ 、 $E_2 = V_2X_2$ 、 $E_3 = V_3X_1^c$ 、 $E_4 = V_4X_2^c$ 、 $E_5 = V_5X_1X_2$ 、 $E_6 = V_6X_1^cX_2$ 、 $E_7 = V_7X_1X_2^c$ 、 $E_8 = V_8X_1^cX_2^c$ 。這些機制造成的結果 Y_V 可以分別解釋為： V_0 代表經由風險因子以外的其他原因使結果發生；經由 V_1X_1 機制使結果 Y_V 發生，其中 X_1 是風險因子；經由 V_2X_2 機制使結果 Y_V 發生，其中 X_2 是風險因子；經由 $V_3X_1^c$ 機制使結果 Y_V 發生，其中 X_1^c 代表風險因子 X_1 為必須成份因，它必須存在但是不作用；經由 $V_4X_2^c$ 機制使結果 Y_V 發生，其中 X_2^c 代表風險因子 X_2 為必須成份因，它必須存在但是不作用；經由 $V_5X_1X_2$ 機制使結果 Y_V 發生，其中風險因子 X_1 和 X_2 必須同時作用才會使機制發生；經由 $V_6X_1^cX_2$ 機制使結果 Y_V 發生，其中風險因子 X_1 為必須成份因，它必須存在但是不作用，但是 X_2 必須作用，兩者都成立才會使機制發生；經由 $V_7X_1X_2^c$ 機制使結果 Y_V 發生，其中風險因子 X_1 必須表現，但是 X_2 為必須成份因，它必須存在但是不作用，也是兩者都成立才會使機制發生；經由 $V_8X_1^cX_2^c$ 機制使結果 Y_V 發生，其中 X_1 和 X_2 都是必須成份因，它必須存在但是不作用。因此結果變數可以由以上所有可能機制事件的，經由析取或運算子以式子來表示：

$$Y_V = V_0 \vee V_1X_1 \vee V_2X_2 \vee V_3X_1^c \vee V_4X_2^c \vee V_5X_1X_2 \vee V_6X_1^cX_2 \vee V_7X_1X_2^c \vee V_8X_1^cX_2^c \quad (\text{式 3.1})$$

因此如果在這裡考慮反事實結果，經由(式 3.1)可以建構有兩個風險因子變數的反事實結果變數

$$Y(x_1, x_2) \leftrightarrow \max\{V_0, V_1x_1, V_2x_2, V_3(1-x_1), V_4(1-x_2), V_5x_1x_2, \\ V_6(1-x_1)x_2, V_7x_1(1-x_2), V_8(1-x_1)(1-x_2)\} \quad (\text{式 3.2})$$

有了在充分原因模型建構出的反事實結果，可以進一步考慮風險因子可能的四種情況， $(x_1, x_2) = (1, 1)$ ， $(x_1, x_2) = (1, 0)$ ， $(x_1, x_2) = (0, 1)$ ， $(x_1, x_2) = (0, 0)$ ，可以把四種反事實結果以向量表示

$$(Y(1, 1), Y(1, 0), Y(0, 1), Y(0, 0)) \quad (\text{式 3.3})$$

反應模式(式 3.3)代表在所有可能的風險因子情況下的反事實結果。其中的符號 $Y(i, j)$ 分別代表 $Y(1, 1) = Y(x_1 = 1, x_2 = 1)$ ， $Y(1, 0) = Y(x_1 = 1, x_2 = 0)$ ， $Y(0, 1) = Y(x_1 = 0, x_2 = 1)$ ， $Y(0, 0) = Y(x_1 = 0, x_2 = 0)$ 。

另一方面當我們把充分條件原因的所有可能機制同時以向量來呈現 $(V_0, V_1, V_2, V_3, V_4, V_5, V_6, V_7, V_8)$ ，並且考慮背景成份因是否完成，並且以二元變數表示，它們總共可以有 $2^9 = 512$ 種組合方式。當我們有了背景成份因的所有組合和兩個風險因子的四種情況，由(式 3.2)可以讓我們知道這512種不同的可能組合方式如何各自對應到(式 3.3)的反應模式。在這 16 種反應模式中， $(0, 0, 0, 0)$ 、 $(0, 0, 0, 1)$ 、 $(0, 1, 0, 0)$ 、 $(0, 0, 1, 0)$ 、 $(0, 1, 1, 0)$ 、 $(1, 0, 0, 0)$ 、 $(1, 0, 0, 1)$ 對應到 1 個背景成份因組合； $(0, 1, 0, 1)$ 、 $(0, 0, 1, 1)$ 、 $(1, 1, 0, 0)$ 、 $(1, 0, 1, 0)$ 、對應到 5 個背景成份因的組合； $(0, 1, 1, 1)$ 、 $(1, 1, 0, 1)$ 、 $(1, 0, 1, 1)$ 、 $(1, 1, 1, 0)$ 對應到 17 個背景成

份因的組合；(1, 1, 1, 1)對應到 417 個背景成份因組合。

從背景成份因的個數來看，對應到一個背景成份因組合的反應模式大部分是一個反事實結果等於 1，它包括了三個拮抗作用(0, 1, 0, 0)、(0, 0, 1, 0)、(0, 1, 1, 0)、兩個協同作用(1, 0, 0, 0)、(1, 0, 0, 1)、一個對風險因子重編碼的協同作用(0, 0, 0, 1)、一個免疫反應模式(0, 0, 0, 0)；對應到 5 個背景成份因組合的反應模式有兩個反事實結果等於 1，它包括了；單一風險因子所呈現的反事實結果(1, 1, 0, 0)、(1, 0, 1, 0)，兩個風險因子的協同作用(0, 1, 0, 1)、(0, 0, 1, 1)；對應 17 個背景成份因組合的反應模式有三個反事實結果等於 1，它包括了促效作用的反應模式(1, 1, 1, 0)、對風險因子重新編碼後呈現促效作用的反應模式(0, 1, 1, 1)、(1, 1, 0, 1)、(1, 0, 1, 1)；對應 417 個背景成份因組合的反應模式有四個反事實結果等於 1 的(1, 1, 1, 1)，它表示不管風風險因子是否呈現，它的背景成份因 V_0 就可以充分使反事實結果發生。

由以上的充分組成原因機制的背景條件向量，對應到反應模式的個數，可以發現在討論結果發生的背後機制的辨別上，考慮兩個的風險因子的反應模式，可以去辨別恰好對應的一個向量的某個機制。當我們從反應模式來看拮抗作用和協同作用，它只會對應到一組在充分成份原因背景條件成份因的組合。因此如果方法可以估計這兩個機制對應的反應模式平均值，就可以量化這兩種邊際機制的大小。另外當考慮促效作用的反應模式時，它所對應背景成份因組合會比拮抗作用和協同作用來的個數多，代表它比另外兩個交互作用機制複雜。

3.2. 考慮 B 型與 C 型肝炎病毒的拮抗機制

3.2.1. 充分成份原因框架下的拮抗機制

上一節我們考慮在風險因子對結果完全沒有預防作用的情況下，充分成份原因如何對應反應模式，這一節考慮肝癌和它的風險因子資料的實際機制。當結果變數沒有單調遞增，一個風險因子對結果變數的影響會有致病機制和預防機制兩種，兩個風險因子的經由交互作用產生的結果變數也會再細分若干分類。

由 Rothman [2] 的對充分成份原因的分類與命名，九種機制分別是不考慮風險因子的 V_0 機制作用，單一風險因子導致結果的 V_1X_1 和 V_2X_2 ，單一風險因子產生預防結果的 $V_3X_1^c$ 和 $V_4X_2^c$ ，協同機制 $V_5X_1X_2$ ，過去文獻上定義的兩個風險因子拮抗機制 $V_6X_1^cX_2$ 和 $V_7X_1X_2^c$ ，重新編碼的協同機制 $V_8X_1^cX_2^c$ ，等共九個機制。結果變數以上機制的勝率運算子呈現(式 3.1) [6, 31]。

本研究方法在探討結果背後發生的機制時，如果過去的文獻研究沒有支持該機制的存在，在模型建立與分析時，不考慮充分成份原因的部分機制。當兩個風險因子分別是 C 型肝炎病毒 X_1 和 B 型肝炎病毒 X_2 ，結果變數是肝癌 Y_V 的情況下，由於這兩類病毒對肝癌是高風險的因子，加上兩者之間的交互作用，充分成份原因的九種機制存在與否必須重新評估。由過去研究的結果顯示這兩種肝炎病毒對肝癌是高風險因子，因此在這裡我們不考慮這風險因子對結果變數的預防機制 $V_3X_1^c$ 和 $V_4X_2^c$ ，也不考慮兩個風險因子同時存在而對結果產生的預防機

制 $V_8X_1^cX_2^c$ 。由於目前的研究並沒有 B 型肝炎病毒會抑制 C 型肝炎病毒的作用，所以 $V_7X_1X_2^c$ 的機制先排除不討論。把以上保留的機制背景條件一起考慮，結果變數可以由析取或運算子表示成 (式 3.4) 的單一風險因子、協同、拮抗等五個機制的聯集。

$$Y_V = V_0 \vee V_1X_1 \vee V_2X_2 \vee V_5X_1X_2 \vee V_6X_1^cX_2 \quad (\text{式 3.4})$$

在這裡其中 V_0 代表肝癌發生的機制不是經由 B 型肝炎病毒和 C 型肝炎病毒這兩個風險因子引起； V_1X_1 代表肝癌發生的機制是經由 C 型肝炎病毒風險因子 X_1 引起； V_2X_2 代表肝癌發生的機制是經由 B 型肝炎病毒風險因子 X_2 引起； $V_5X_1X_2$ 代表肝癌發生的機制是經由 C 型肝炎病毒和 B 型肝炎病毒這兩個風險因子的協同作用引起； $V_6X_1^cX_2$ 代表肝癌發生的機制是經由 C 型肝炎病毒和 B 型肝炎病毒這兩個風險因子的拮抗作用引起。

3.2.2. 反應模式與邊際充分成份原因

與 3.1 節一樣的方法，我們把三十二種由二元變數組成的五個背景成份因 (V_0, V_1, V_2, V_5, V_6) 和兩個二元風險因子的四種組合帶入 (式 3.4)，然後依照 (式 3.3) 反應模式把這些背景成份因組合做歸類。

藉由比較反應模式和充分成份原因 (式 3.4)，充分成份原因可以被經由合併得到邊際的充分成份原因。考慮在介入兩個風險因子，由四種反事實結果所形成

的反應模式和充分成份原因的背景條件，經由(式 3.2)我們可以把三十二種背景情況歸類到七種反應模式。其中反應模式(0, 0, 0, 0)、(0, 0, 1, 0)、(1, 0, 0, 0)對應的一個成份因組合；(1, 1, 0, 0)對應到兩個成份因組合；(1, 0, 1, 0)對應到五個成份因組合；(1, 1, 1, 0)對應到六個成份因組合；(1, 1, 1, 1)對應到十六個成份因組合。我們發現只有拮抗機制 V_6 呈現的成份因組合只對應到一個反應模式；同樣的只有協同機制 V_5 呈現的成份因組合也只對應到一個反應模式。其它的成份因組合都至少有兩個以上對應到同一個反應模式。

表格 2：在充分成份原因模型，考慮拮抗、協同兩種機制交互作用存在的情況下，背景成分組成的向量與反應模式的對應關係。

Response patterns ($Y(1, 1), Y(1, 0), Y(0, 1), Y(0, 0)$)	Background condition vector (V_0, V_1, V_2, V_5, V_6)
(1, 1, 1, 1)	(1, 0/1, 0/1, 0/1, 0/1) (There are sixteen combinations)
(1, 1, 0, 0)	(0, 1, 0, 0, 0), (0, 1, 0, 1, 0)
(1, 0, 1, 0)	(0, 0, 0, 1, 1), (0, 0, 1, 0, 0), (0, 0, 1, 0, 1), (0, 0, 1, 1, 0), (0, 0, 1, 1, 1)
(1, 0, 0, 0)	(0, 0, 0, 1, 0)
(0, 0, 1, 0)	(0, 0, 0, 0, 1)
(1, 1, 1, 0)	(0, 1, 1, 1, 1), (0, 1, 1, 1, 0), (0, 1, 1, 0, 1), (0, 1, 1, 0, 0), (0, 1, 0, 1, 1), (0, 1, 0, 0, 1)
(0, 0, 0, 0)	(0, 0, 0, 0, 0)

我們依據這七種反應模式所對應的背景成份因組合(V_0, V_1, V_2, V_5, V_6)同時出

現與否，以集合的運算方式表達，在同一括號的背景條件用交集運算，在不同括號的背景條件用聯集運算，來歸納對應的邊際效果。

I. 在這樣的集合運算下，反應模式(1, 1, 1, 1)的成份因組合可以表示成

V_0, V_1, V_2, V_5, V_6 和 $V_1^c, V_2^c, V_5^c, V_6^c$ 的聯集，因此可用 V_0 代表它。

II. 反應模式(1, 1, 0, 0)的成份因組合為 $(V_0^c \cap V_1 \cap V_2^c \cap V_5^c \cap V_6^c) \cup (V_0^c \cap V_1 \cap V_2^c \cap V_5 \cap V_6^c) = V_0^c \cap V_1 \cap V_2^c \cap V_6^c$ ，因此可用 V_1 代表它。

III. 反應模式(1, 0, 1, 0)的成份因組合為 $(V_0^c \cap V_1^c \cap V_2^c \cap V_5 \cap V_6) \cup (V_0^c \cap V_1^c \cap V_2 \cap V_5^c \cap V_6^c) \cup (V_0^c \cap V_1^c \cap V_2 \cap V_5 \cap V_6^c) \cup (V_0^c \cap V_1^c \cap V_2 \cap V_5 \cap V_6) = (V_0^c \cap V_1^c \cap V_2^c \cap V_5 \cap V_6) \cup (V_0^c \cap V_1^c \cap V_2)$ ，由於背景成份因 V_5, V_6 的風險因子都有 X_2 ，因此在這以 V_2 代表反應模式(1, 0, 1, 0)。

IV. 反應模式(1, 0, 0, 0)只對應一個成份因組合 $V_0^c \cap V_1^c \cap V_2^c \cap V_5 \cap V_6^c$ ，因此可用 V_5 代表它。

V. 反應模式(0, 0, 1, 0)只對應一個成份因組合 $V_0^c \cap V_1^c \cap V_2^c \cap V_5^c \cap V_6$ ，因此可用 V_6 代表它。

VI. 反應模式(1, 1, 1, 0)對應了六個成份因組合，他們的聯集會是 $(V_0^c \cap V_1 \cap V_2) \cup (V_0^c \cap V_2^c \cap V_1 \cap V_6)$ ，在這個情況無法選擇對應的背景成份因，由於它們對應到不相同的風險因子。

VII. 反應模式(0,0,0,0)對應了不影響結果的背景條件交集 $V_0^c \cap V_1^c \cap V_2^c \cap V_5^c \cap V_6^c$ 。

如果我們從這些反應模式觀察上述 I-VII 集合運算後其中不是補集的背景成份因，我們可以把這些反應模式分別歸類為

- I. (1, 1, 1, 1)結果發生和風險因子無關的模式
- II. (1, 1, 0, 0)一個風險因子 X_1 邊際效果
- III. (1, 0, 1, 0)一個風險因子 X_2 邊際效果
- IV. (1, 0, 0, 0)兩個風險因子 X_1 和 X_2 協同機制邊際效果
- V. (0, 0, 1, 0)兩個風險因子 X_1 和 X_2 拮抗機制邊際效果
- VI. (1, 1, 1, 0)兩個風險因子 X_1 和 X_2 促效機制邊際效果
- VII. (0, 0, 0, 0)風險因子無關不會影響結果的反應模式

表格 3：由充分成份原因的背景條件，考慮背景條件的存在與不存在，以集合的非補集與補集表示(欄位左 2)，然後對應到有關的風險因子(欄位右 2)。最後歸納出反應模式(欄位左 1)與邊際成分原因(欄位右 1)。

Response patterns ($Y(1, 1), Y(1, 0), Y(0, 1), Y(0, 0)$)	Collected sufficient background components (V_0, V_1, V_2, V_5, V_6)	Marginal Risk factors	Marginal sufficient component causes
(1, 1, 1, 1)	V_0	None	U_0

(1, 1, 0, 0)	$V_0^c \cap V_2^c \cap V_6^c \cap V_1$	X_1	$U_1 X_1$
(1, 0, 1, 0)	$(V_0^c \cap V_1^c \cap V_2^c \cap V_5 \cap V_6) \cup (V_0^c \cap V_1^c \cap V_2)$	X_2	$U_2 X_2$
(1, 0, 0, 0)	$V_0^c \cap V_1^c \cap V_2^c \cap V_6^c \cap V_5$	$X_1 X_2$	$U_3 X_1 X_2$
(0, 0, 1, 0)	$V_0^c \cap V_1^c \cap V_2^c \cap V_5^c \cap V_6$	$X_1^c X_2$	$U_4 X_1^c X_2$
(1, 1, 1, 0)	$(V_0^c \cap V_1 \cap V_2) \cup (V_0^c \cap V_2^c \cap V_1 \cap V_6)$	X_1 or X_2	$U_5 (X_1 + X_2 - X_1 X_2)$
(0, 0, 0, 0)	$V_0^c \cap V_1^c \cap V_2^c \cap V_5^c \cap V_6^c$	None	None

在 I-VII 當中值得提及的是拮抗機制的反應模式 $(Y(1, 1), Y(1, 0), Y(0, 1), Y(0, 0)) = (0, 0, 1, 0)$ 恰好對應到一組 SCC 的背景條件 $(V_0, V_1, V_2, V_5, V_6) = (0, 0, 0, 0, 1)$ 。所以當我們由該反應模式是否呈現，來堆論反映結果的發生是否背後存在是單獨經由拮抗機制所引起。同樣的情況下，由於反應模式 $(0, 0, 1, 0)$ 恰好對應到一組 SCC 的背景條件 $(V_0, V_1, V_2, V_5, V_6) = (0, 0, 0, 1, 0)$ 。所以當我們也可以堆論反應結果的發生是否背後存在是單獨經由協同機制所引起。

3.2.3. mSCC 的背景條件與 SCC 背景條件的對應

上一小節在說明了反應模式和背景成份因組成的向量的對應，然後得到風險因子的邊際效果。由於經由(式 3.4)的轉換把反是實結果的反應模式和充分成份原因的機制連結，在原本的充分成份原因，它的背景成份因組成的向量也隨著

對應到了反應模式，所以風險因子的邊際機制，也會同時一起對應到了這些背景成份因組成的向量。它們被稱作邊際背景成份因 (background component of mSCC)。在此我們會由 3.2.2 的結果來說明這些邊際背景成份因。

I. $U_0 = V_0$

U_0 表示和風險因子無關的背景成份因。在充分成份原因裡面，它剛好等於和風險因子無關的機制。

II. $U_1 = V_0^c \cap V_1 \cap V_2^c \cap V_6^c$

U_1 表示和風險因子 X_1 邊際背景成份因。在充分成份原因裡面，它對應到了風險因子 X_1 背景成份因 V_1 ，和兩個風險因子協同機制的 V_5 ，同時必須滿足其它機制的背景成份因不存在。

III. $U_2 = (V_0^c \cap V_1^c \cap V_2^c \cap V_5 \cap V_6) \cup (V_0^c \cap V_1^c \cap V_2)$

U_2 表示和風險因子 X_2 邊際背景成份因。在充分成份原因裡面，它對應到了風險因子 X_2 背景成份因 V_2 ，但是 V_0 和 V_1 這兩個背景成份因不存在；或是和兩個風險因子的協同機制與拮抗機制的背景成份因 V_5 與 V_7 同時存在，但是 V_0 、 V_1 和 V_2 這三個背景成份因不存在。

IV. $U_3 = V_0^c \cap V_1^c \cap V_2^c \cap V_5 \cap V_6^c$

U_3 表示和風險因子 X_1 和 X_2 邊際協同機制的背景成份因。在充分成份原因裡面，它對應到了 X_1 和 X_2 背景成份因 V_5 ，但是 V_0 、 V_1 、 V_2 和 V_6 這四個背景成份因不存在。

V. $U_4 = V_0^c \cap V_1^c \cap V_2^c \cap V_5^c \cap V_6$

U_4 表示和風險因子 X_1 和 X_2 邊際拮抗機制的背景成份因。在充分成份原因裡面，它對應到了 X_1 和 X_2 背景成份因 V_6 ，但是 V_0 、 V_1 、 V_2 和 V_5 這四個背景成份因不存在。

VI. $U_5 = (V_0^c \cap V_1 \cap V_2) \cup (V_0^c \cap V_2^c \cap V_1 \cap V_6)$

U_5 表示和風險因子 X_1 和 X_2 促效機制的背景成份因。在充分成份原因裡面 V_0 不存在的前提下不考慮 V_3 存在與否，它對應到了 X_1 和 X_2 背景成份因 V_1 和 V_2 ；或是 V_1 和 V_6 其中的兩個背景成份因同時存在的情況。

$$\text{VII. } U_{-1} = V_0^c \cap V_1^c \cap V_2^c \cap V_5^c \cap V_6^c$$

U_{-1} 表示在充分成份原因裡面所有的背景成份因都必須不存在。

3.3. 邊際充分成份原因模型的拮抗機制

在上面兩節當中把風險因子的機制經由(式 3.4)對應到(式 3.2)的反應模式後，就可以經由反事實結果的概念，從觀察的資料來估計，來量化個別機制影響結果變數的比例。在這裡我們將依序對邊際充分成份原因建立模型，說明模型的反事實結果形式，然後辨別模型裡的各機制，最後用迴歸模型給予各機制的估計式。

3.3.1. 對邊際充分成份原因的機率模型

要量化邊際充分成份原因機制的大小，必須考慮風險因子和邊際背景成份因的機率。假設二元結果變數 $Y_U = 1$ 的時候只會對應的一個邊際機制，也就是說當 Y_U 的結果發生，不會同時有兩個邊際機制是導致它發生的原因。因此 Y_U 可以經由式子表示成所有可能機制的加總。

$$Y = S_0 + S_1 + S_2 + S_3 + S_4 + S_5 \quad (\text{式 3.5})$$

其中 S_0 是結果發生，但是和風險因子無關的機制。 S_1 是結果發生，邊際背景

成份因和因子 X_1 有關的機制。 S_2 是結果發生，邊際背景成份因和因子 X_2 有關的機制。 S_3 是結果發生，邊際背景成份因和因子 X_1 、 X_2 有關的邊際協同機制。 S_4 是結果發生，邊際背景成份因和因子 X_1 、 X_2 有關的邊際拮抗機制。 S_5 是結果發生，邊際背景成份因和因子 X_1 、 X_2 有關的邊際促效機制。

為了描述個機制的發生機率，可以進一步假設 U 是一個類別隨機變數，它代表 $Y = 1$ 所對應的發生機制的邊際背景成份因，類別隨機變數 U 的範圍可以表示成 $U \in \{-1, 0, 1, 2, 3, 4, 5\}$ 。假設當 $Y = 1$ 發生時只會對應其中某一個機制，而它只會對應到某一邊際背景成份向量的類別，因此我們再對各邊際機制的邊際背景成份的發生與否，定義指標隨機變數 $U_j = I(U = j)$ ，其中 $j = -1, 0, 1, 2, 3, 4, 5$ 。因此結果變數 Y_U 可以進一步由兩個風險因子和邊際背景成份因呈現。

$$Y = U_0 + U_1X_1 + U_2X_2 + U_3X_1X_2 + U_4(1 - X_1)X_2 + U_5(X_1 + X_2 - X_1X_2) \quad (式 3.6)$$

由式子 3.6 把結果變數由因子和背景成份因呈現，可以考慮個體的反事實結果

$$Y(x_1, x_2) = U_0 + U_1x_1 + U_2x_2 + U_3x_1x_2 + U_4(1 - x_1)x_2 + U_5(x_1 + x_2 - x_1x_2) \quad (式 3.7)$$

其中(式 3.7)的 $U_4(1 - x_1)x_2$ 、 $U_5(x_1 + x_2 - x_1x_2)$ 、 $U_3x_1x_2$ 分別代表反事實結果

在個體尺度下的邊際拮抗機制、邊際促效機制、邊際協同機制。群體尺度的反事實

實結果可以由對(式 3.7)取期望值表示。

$$\begin{aligned}
 E[Y(x_1, x_2)] &= E[U_0] + E[U_1x_1] + E[U_2x_2] + E[U_3x_1x_2] + E[U_4(1-x_1)x_2] \\
 &\quad + E[U_5(x_1 + x_2 - x_1x_2)] \\
 &= E[U_0] + E[U_1]x_1 + E[U_2]x_2 + E[U_3]x_1x_2 + E[U_4](1-x_1)x_2 \\
 &\quad + E[U_5](x_1 + x_2 - x_1x_2) \\
 &= p_{U_0} + p_{U_1}x_1 + p_{U_2}x_2 + p_{U_3}x_1x_2 + p_{U_4}(1-x_1)x_2 + p_{U_5}(x_1 + x_2 - x_1x_2)
 \end{aligned}$$

$$\begin{aligned}
 E[Y(x_1, x_2)] &= p_{U_0} + p_{U_1}x_1 + p_{U_2}x_2 + p_{U_3}x_1x_2 + p_{U_4}(1-x_1)x_2 \\
 &\quad + p_{U_5}(x_1 + x_2 - x_1x_2) \tag{式 3.8}
 \end{aligned}$$

其中(式 3.8)的 p_{U_4} 、 p_{U_5} 、 p_{U_3} 分別代表反事實結果在群體尺度下的邊際拮抗機制、邊際促效機制、邊際協同機制的三個比例。 p_{U_0} 、 p_{U_1} 、 p_{U_2} 、 p_{U_3} 、 p_{U_4} 、 p_{U_5} 分別代表由指標變數定義的每個邊際機制背景成份因的機率， $\Pr(U_j = 1) = p_{U_j}$ 。

由於 Y_U 的發生必定是來自於其中的一個機制，因此它們的機率和等於1。

$$\sum_{j=-1}^4 p_{U_j} = p_{U_{-1}} + p_{U_0} + p_{U_1} + p_{U_2} + p_{U_3} + p_{U_4} + p_{U_5} = 1 \tag{式 3.9}$$

然而這裡的目標是估計六個機制的大小，然而(式 3.8)的自變數只有兩個，線性

模式的飽和模型只能估計四個機制。

$$\begin{aligned} E[Y(x_1, x_2)] &= p_{U_0} + (p_{U_1} + p_{U_5})x_1 + (p_{U_2} - p_{U_4} + p_{U_5})x_2 \\ &+ (p_{U_3} - p_{U_4} + p_{U_5})x_1x_2 \end{aligned} \quad (\text{式 } 3.10)$$

從(式 3.10)可以發現如果使用線性飽和模型來估計，拮抗機制無法經由 x_1 、 x_2 、 x_1x_2 的係數方程式求得。因此在這裡需要增加模型自由度來辨別拮抗機制。為了辨識邊際拮抗機制與邊際促效機制，我們將會在模型中加入兩個代理變數。

3.3.2. 代理變數的定義和性質

根據邊際充分原因模型的研究文獻來歸納[9, 10]，依據對因子作用機制的瞭解程度，對代理變數分別做以下兩個假設。

代理變數必須目標機制的背景條件的一個組成元素。 (假設 1)

在具備足夠的知識而瞭解多個因子與結果變數的機制之下，經由確認代理變數不是導致結果的風險因子之後，可以從風險因子以外當中的元素，由假設 1 來選擇而外的元素或因子做為代理變數，來辨識想估計的機制。例如生物學領域在細胞尺度下的機制，在可以確定代理變數可以是在機制導致結果發生的過程中，接收子(receptor)所對應的生物標記(biomarker)，這時的生物標記可以作為假設 1 的代理變數。另一個對代理變數的假設為

代理變數是對結果造成影響的其它風險因子。而且它與原本的風險因子彼此之間也會透過交互作用對結果產生影響。 (假設 2)

在對於其它因子對於結果影響的瞭解有限之下，為了明確辨識想估計的機制，假設 2 把代理變數視為可能的風險因子，考慮與所有風險因子之間的交互作用。

這裡考慮滿足假設 1 的代理變數條件，分別從邊際拮抗與邊際促效兩個機制的邊際背景成份因 U_4 、 U_5 ，各自確認並選擇額外的因子來分別做為兩個機制的代理變數。從拮抗機制的邊際背景成份因 U_4 挑選出因子 Z_1 ，它滿足假設 1。換句話說，因子 Z_1 是在因子 X_1 、 X_2 的拮抗機制交互作用過程裡，其它的必要元素之一。因子 Z_1 與 X_1 、 X_2 不同的地方在於它不是結果的風險因子。以 U_{40} 代表 U_4 扣掉 Z_1 之後剩餘的背景成份因，拮抗機制的背景成份 U_4 可以重新表示成 $U_4 = U_{40}Z_1$ 。

同樣的方法，從促效機制的邊際背景成份因 U_5 挑選出因子 Z_2 ，它也是滿足假設 1。因子 Z_2 不是 X_1 、 X_2 促效機制的風險因子。以 U_{50} 代表 U_5 扣掉 Z_2 之後剩餘的背景成份因，促效機制的背景成份 U_5 可以重新表示成 $U_5 = U_{50}Z_2$ 。因此給定兩個因子 X_1 、 X_2 和兩個代理變數 Z_1 、 Z_2 之下，個體的反事實結果 Y_U 為

$$\begin{aligned} Y(x_1, x_2, z_1, z_2) &= U_0 + U_1x_1 + U_2x_2 + U_3x_1x_2 + U_{40}z_1(1 - x_1)x_2 \\ &\quad + U_{50}z_2(x_1 + x_2 - x_1x_2) \end{aligned} \tag{式 3.11}$$

對(式 3.11)取期望值，可以得到兩個因子和兩個代理變數的群體反事實結果

$$\begin{aligned}
& E[Y(x_1, x_2, z_1, z_2)] \\
&= E[U_0] + E[U_1x_1] + E[U_2x_2] + E[U_3x_1x_2] + E[U_4(1-x_1)x_2] \\
&+ E[U_5(x_1 + x_2 - x_1x_2)] \\
&= E[U_0] + E[U_1]x_1 + E[U_2]x_2 + E[U_3]x_1x_2 + E[U_4](1-x_1)x_2 \\
&+ E[U_5](x_1 + x_2 - x_1x_2) \\
&= E[U_0] + E[U_1]x_1 + E[U_2]x_2 + E[U_3]x_1x_2 + E[U_{40}]z_1(1-x_1)x_2 \\
&+ E[U_{50}]z_2(x_1 + x_2 - x_1x_2) \\
&= p_{U_0} + p_{U_1}x_1 + p_{U_2}x_2 + p_{U_3}x_1x_2 + p_{U_{40}}z_1(1-x_1)x_2 \\
&+ p_{U_{50}}z_2(x_1 + x_2 - x_1x_2) \\
&E[Y(x_1, x_2, z_1, z_2)] \\
&= p_{U_0} + p_{U_1}x_1 + p_{U_2}x_2 + p_{U_3}x_1x_2 + p_{U_{40}}z_1(1-x_1)x_2 \\
&+ p_{U_{50}}z_2(x_1 + x_2 - x_1x_2) \tag{式 3.12}
\end{aligned}$$

經由邊際充分成份原因模型建立了群體反事實結果，可以借助因果推論與統計模型方法，由資料來估計(式 3.12)的各個機制。

3.4. 辨識拮抗作用的機制

考慮結果變數 Y 受到因子 X_1 、 X_2 、 U 的影響，其中因子 X_1 、 X_2 、 U 受到因子 C 的影響(圖 1)。

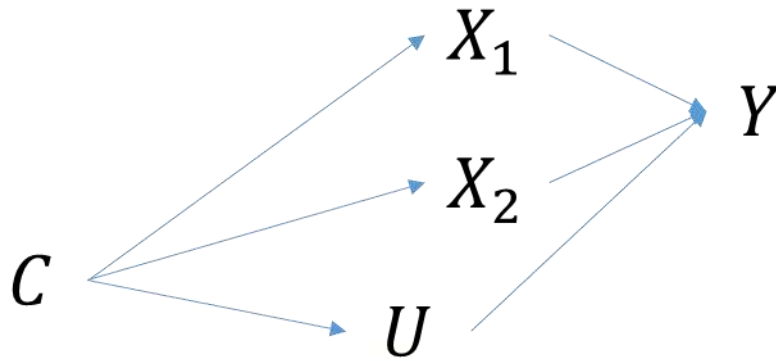


圖 1：兩個因子，背景條件因子，邊際結果，干擾因子等四者之間的因果非有向循環圖。

在給定因子 $C = c$ 的分層之下，因子 X_1 、 X_2 、 U 各自對結果 Y_U 的因果關係彼此之間不互相影響，因此可以得到 $E[Y_U(x_1, x_2)] = E_C[E[Y_U(x_1, x_2)|C]]$ 。根據因果推論的可交換條件(exchangeability)假設：

$$Y(x_1, x_2) \perp X_1 | C \quad (\text{假設 3})$$

$$Y(x_1, x_2) \perp X_2 | C \quad (\text{假設 4})$$

反事實結果期望值可以表示成給定兩個因子已知情況下的反事實結果期望值。

$$E[Y(x_1, x_2)|C] = E[Y(x_1, x_2)|X_1 = x_1, X_2 = x_2, C] \quad (\text{式 3.13})$$

再由因果推論的一致性(consistency)假設：

$$\text{給定 } X_1 = x_1, X_2 = x_2, C, Y(x_1, x_2) \text{ 與 } Y \text{ 相等。} \quad (\text{假設 5})$$

兩個因子已知情況下的反事實結果期望值，可以再被表示成觀察到結果的條件

期望值。

$$E[Y(x_1, x_2)|X_1 = x_1, X_2 = x_2, C] = E[Y|X_1 = x_1, X_2 = x_2, C] \quad (\text{式 } 3.14)$$

然而由兩個自變數的線性飽和模型的自由度不足以估計(式 3.14)的六個機制。在這裡我們分別引入 U_4 機制的代理變數 Z_1 、 U_5 機制的代理變數 Z_2 ，並且假設從機制的背景成份因獨立出的代理變數不會受到因子 C 的影響(圖 2)。

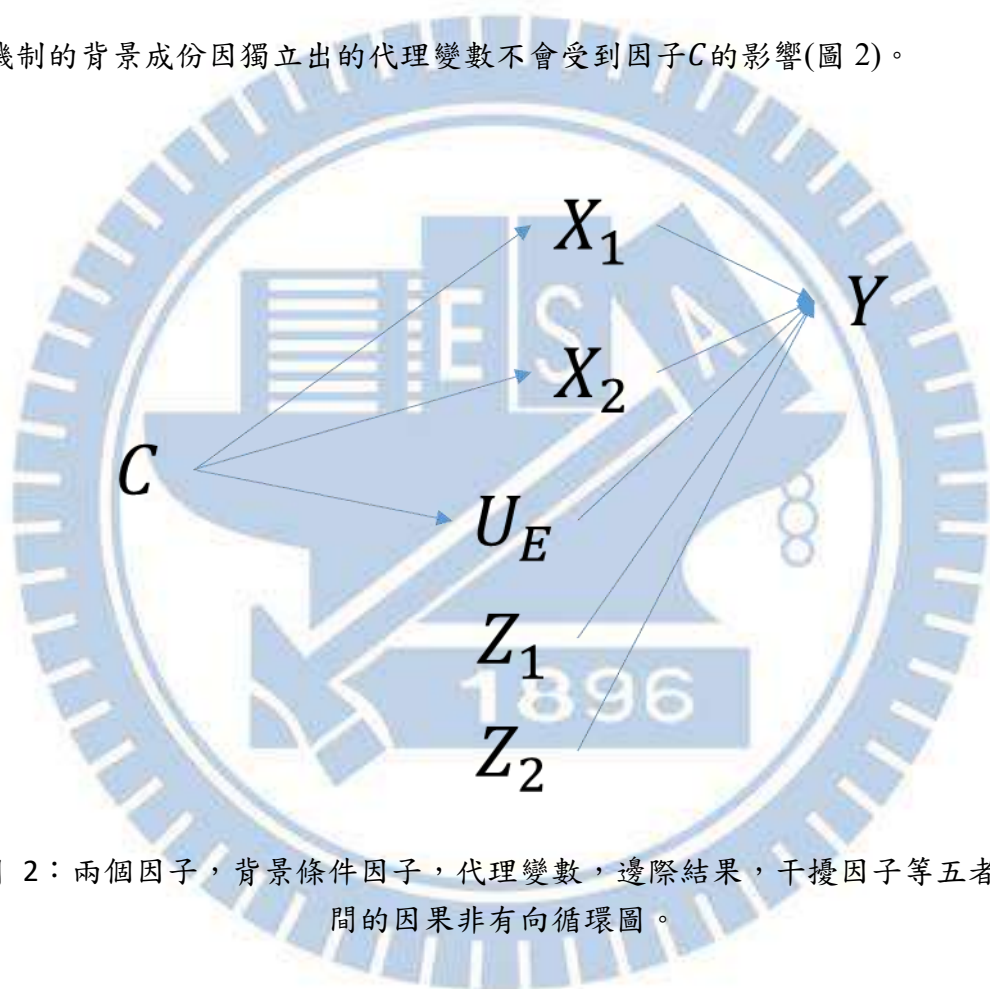


圖 2：兩個因子，背景條件因子，代理變數，邊際結果，干擾因子等五者之間的因果非有向循環圖。

在給定因子 $C = c$ 的分層之下，因子 X_1 、 X_2 、 U_E 、 Z_1 、 Z_2 各自對結果 Y_U 的因果關係彼此之間不互相影響，因此可以得到 $E[Y(x_1, x_2, z_1, z_2)] = E_C[E[Y(x_1, x_2, z_1, z_2)|C]]$ 。根據因果推論的可交換條件假設 3、4、6、7，

$$Y(x_1, x_2, z_1, z_2) \perp Z_1 \quad (\text{假設 } 6)$$

$$Y(x_1, x_2, z_1, z_2) \perp Z_2 \quad (\text{假設 7})$$

反事實結果期望值可以表示成給定兩個因子已知情況下的反事實結果期望值。

$$\begin{aligned} E[Y(x_1, x_2, z_1, z_2)] &= E[Y(x_1, x_2, z_1, z_2)|Z_1 = z_1, Z_2 = z_2] \\ &= E_C[E[Y(x_1, x_2, z_1, z_2)|Z_1 = z_1, Z_2 = z_2, C]] \\ &= E_C[E[Y(x_1, x_2, z_1, z_2)|X_1 = x_1, X_2 = x_2, Z_1 = z_1, Z_2 = z_2, C]] \quad (\text{式 3.15}) \end{aligned}$$

再由因果推論的一致性假設，

$$\text{給定 } X_1 = x_1, X_2 = x_2, Z_1 = z_1, Z_2 = z_2, C, Y(x_1, x_2, z_1, z_2) \text{ 與 } Y \text{ 相等。} \quad (\text{假設 8})$$

兩個因子與兩個代理變數都是已知情況下的反事實結果期望值，可以再被表示成觀察到結果的條件期望值。

$$\begin{aligned} E[Y(x_1, x_2, z_1, z_2)|X_1 = x_1, X_2 = x_2, Z_1 = z_1, Z_2 = z_2, C] \\ = E[Y|X_1 = x_1, X_2 = x_2, Z_1 = z_1, Z_2 = z_2, C] \quad (\text{式 3.16}) \end{aligned}$$

藉由(式 3.16)我們便可以由統計模型與觀察的資料來同時估計邊際拮抗機制、邊際促效機制、邊際協同機制這三種交互作用。

3.5. 估計

假設(式 3.16)的線性迴歸模型為

$$\begin{aligned}
& E[Y|X_1 = x_1, X_2 = x_2, Z_1 = z_1, Z_2 = z_2, C = c] \\
&= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_{40} z_1 (1 - x_1) x_2 \\
&+ \beta_{50} z_2 (x_1 + x_2 - x_1 x_2) + \beta_{c0} c + \beta_{c1} c x_1 + \beta_{c2} c x_2 + \beta_{c3} c x_1 x_2 \\
&+ \beta_{c40} c z_1 (1 - x_1) x_2 + \beta_{c50} c z_2 (x_1 + x_2 - x_1 x_2) \tag{式 3.17}
\end{aligned}$$

$$\begin{aligned}
& E_C[E[Y(x_1, x_2, z_1, z_2)|X_1 = x_1, X_2 = x_2, Z_1 = z_1, Z_2 = z_2, C]] \\
&= E_C[\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_{40} z_1 (1 - x_1) x_2 \\
&+ \beta_{50} z_2 (x_1 + x_2 - x_1 x_2) + \beta_{c0} C + \beta_{c1} C x_1 + \beta_{c2} C x_2 + \beta_{c3} C x_1 x_2 \\
&+ \beta_{c40} C z_1 (1 - x_1) x_2 + \beta_{c50} C z_2 (x_1 + x_2 - x_1 x_2)] \\
&= \beta_0 + \beta_{c0} E[C] + (\beta_1 + \beta_{c1} E[C]) x_1 + (\beta_2 + \beta_{c2} E[C]) x_2 \\
&+ (\beta_3 + \beta_{c3} E[C]) x_1 x_2 + (\beta_{40} + \beta_{c40} E[C]) z_1 (1 - x_1) x_2 \\
&+ (\beta_{50} + \beta_{c50} E[C]) z_2 (x_1 + x_2 - x_1 x_2) \tag{式 3.18}
\end{aligned}$$

比較(式 3.12)與(式 3.18)兩個式子的形式，因此我們可以知道

拮抗機制的條件機率估計 $(\hat{\beta}_{40} + \hat{\beta}_{c40} \bar{C}) \bar{Z}_1 (1 - \bar{X}_1) \bar{X}_2$

協同機制的條件機率估計 $(\hat{\beta}_3 + \hat{\beta}_{c3} \bar{C}) \bar{X}_1 \bar{X}_2$

促效機制的條件機率估計 $(\hat{\beta}_{50} + \hat{\beta}_{c50} \bar{C}) \bar{Z}_2 (\bar{X}_1 + \bar{X}_2 - \bar{X}_1 \bar{X}_2)$

以下表格分別呈現在考慮拮抗機制與不考拮抗機制下，對應的邊際成份原因模型的各個機制估計。

表格 4：在邊際成分原因模型下考慮拮抗、協同兩種機制交互作用，對應到在邊際成份原因模型下的拮抗、協同、促效三種邊際機制交互作用。由左而右分別是反應模式、邊際效果的因子、mSCC 的背景成分、機制的理論值、機制的估計值。

Response Patterns	mSCC	Marginal background component	Expected mSCC mechanism	Estimator for mSCC mechanism
(1, 1, 1, 1)	<i>Other components</i>	U_0	$E[U_0 C]$	$\hat{\beta}_0 + \hat{\beta}_{c0}\bar{C}$
(1, 1, 0, 0)	X_1 only	U_1	$E[U_1 C]E[X_1]$	$(\hat{\beta}_1 + \hat{\beta}_{c1}\bar{C})\bar{X}_1$
(1, 0, 1, 0)	X_2 only	U_2	$E[U_2 C]E[X_2]$	$(\hat{\beta}_2 + \hat{\beta}_{c2}\bar{C})\bar{X}_2$
(1, 0, 0, 0)	X_1 and X_2 Synergism	U_3	$E[U_3 C]E[X_1X_2]$	$(\hat{\beta}_3 + \hat{\beta}_{c3}\bar{C})\bar{X}_1\bar{X}_2$
(0, 0, 1, 0)	X_1 and X_2 Antagonism	$U_{40}Z_1$	$E[U_{40} C]E[Z_1]E[(1 - X_1)X_2]$	$(\hat{\beta}_{40} + \hat{\beta}_{c40}\bar{C})\bar{Z}_1(1 - \bar{X}_1)\bar{X}_2$
(1, 1, 1, 0)	X_1 and X_2 Agonsim	$U_{50}Z_2$	$E[U_{50} C]E[Z_2]E[X_1 + X_2 - X_1X_2]$	$(\hat{\beta}_{50} + \hat{\beta}_{c50}\bar{C})\bar{Z}_2(\bar{X}_1 + \bar{X}_2 - \bar{X}_1\bar{X}_2)$

4. 數值分析

4.1. 模擬分析

需要經由模擬產生的變數包括反映結果 Y 、風險因子 X_1 與 X_2 、代理變數 Z_1 與 Z_2 、干擾因子 C 。 Y 的背景條件 U 經由多項分佈的隨機變數產生， $U \sim \text{multinomial}(n, p_{U_0}, p_{U_1}, p_{U_2}, p_{U_3}, p_{U_{40}}, p_{U_{50}}, p_{U_{-1}})$ 。其中 $n = 1$ 。代表 Y 發生時，只對應到一個邊際充分成份原因。參數 p_{U_0} 的機率對應的機制背景條件與因子 X_1 、 X_2 無關。 p_{U_1} 的機率對應的機制背景條件只與風險因子 X_1 有關。 p_{U_2} 的機率對應的機制背景條件只與風險因子 X_2 有關。 p_{U_3} 的機率對應的機制背景條件與風險因子 X_1 、 X_2 有關的協同作用，兩個因子必須都存在於機制作用過程，而且必須有作用。 $p_{U_{40}}$ 的機率對應的機制背景條件與風險因子 X_1 、 X_2 有關的拮抗作用，其中因子 X_1 必須存在但是不作用，因子 X_2 必須存在且作用。 $p_{U_{50}}$ 的機率對應的機制背景條件與風險因子 X_1 、 X_2 有關的促效作用。 $p_{U_{-1}}$ 的機率對應了結果 Y_U 不會發生的機制。為了模擬二元變數的其它因子，干擾因子 C 、風險因子 X_1 、 X_2 、代理變數 Z_1 、 Z_2 等，由白努力分佈的隨機變數來產生。

在模擬資料大小 $N = 1999$ ，模擬次數 $B = 199$ ，每次的模擬可以得到每個邊際成份原因的估計值、誤差、信賴區間以及覆蓋率。六種情境下在模擬資料產生時的參數設定分別如以下所描述。

4.1.1. 考慮拮抗、促效、協同三種機制

成份原因的機率參數維持上述不變的設定。模擬的結果顯示，不經由任何風險因子造成結果Y發生的機率有 0.197 (0.164, 0.230)。由單一風險因子 X_1 的邊際機制造成結果發生的機率有 0.083 (0.057, 0.109)。由單一風險因子 X_2 的邊際機制造成結果發生的機率有 0.146 (0.071, 0.222)。由兩個風險因子 X_1 、 X_2 的邊際協同機制造成結果發生的機率有 0.012 (-0.013, 0.038)。由兩個風險因子 X_1 、 X_2 ，經由代理變數 Z_1 辨識的邊際拮抗機制，這個拮抗機制是因子 X_1 抑制 X_2 的作用，其造成結果發生的機率有 0.033 (-0.011, 0.077)。由兩個風險因子 X_1 、 X_2 ，經由代理變數 Z_2 辨識的邊際促效機制，這個機制是因子 X_1 與因子 X_2 競爭結果Y的發生，其造成結果發生的機率有 0.043 (0.008, 0.078)。

從表格 5 可以發現，兩個因子的主效用與拮抗、促效、協同邊際機制交互作用，它們分別由機率模型參數所計算的理論值，與邊際成份原因模型估計值，兩者之間的差異(bias)都相當接近 0。考慮模擬當中每一次重複估計的信賴區間，這六個機制的平均覆蓋率都接近 95%。因此圖 3 的偏誤值信賴區間有 95%的信心水準會包含 0。這代表如果觀察到的資料如果在充分成份原因模型下有拮抗、促效、協同三種機制存在，可以由邊際成份原因模型估計這三種邊際機制交互作用的大小。換句話說如果在細胞的尺度下有拮抗、促效、協同三種機制存在，則可以經由邊際成份原因模型估計在個體層級的三種邊際機制的大小。

表格 5: 拮抗、促效、協同三種機制交互作用同時存在的模擬結果。表格欄位由左而右分別是理論機率值、比例、估計機率值、偏差值、估計標準差、95%信賴區間上下界、模擬覆蓋率。

	Prob.	Prop.	est. prob.	bias	SE	LCI	UCI	coverage
mechanism U_0	0.194	37.7%	0.197	-0.003	0.017	-0.036	1.974	0.950
mechanism U_1X_1	0.085	16.6%	0.083	0.002	0.013	-0.024	1.975	0.980
mechanism U_2X_2	0.149	28.9%	0.146	0.002	0.038	-0.073	2.001	0.950
synergism $U_3X_1X_2$	0.011	2.2%	0.012	-0.001	0.013	-0.026	1.972	0.945
antagonism $U_{40}Z_1X_1^cX_2$	0.033	6.4%	0.033	0.000	0.023	-0.044	1.982	0.950
agonism $U_{50}Z_2(X_1 + X_2 - X_1X_2)$	0.043	8.3%	0.043	0.000	0.018	-0.035	1.978	0.960

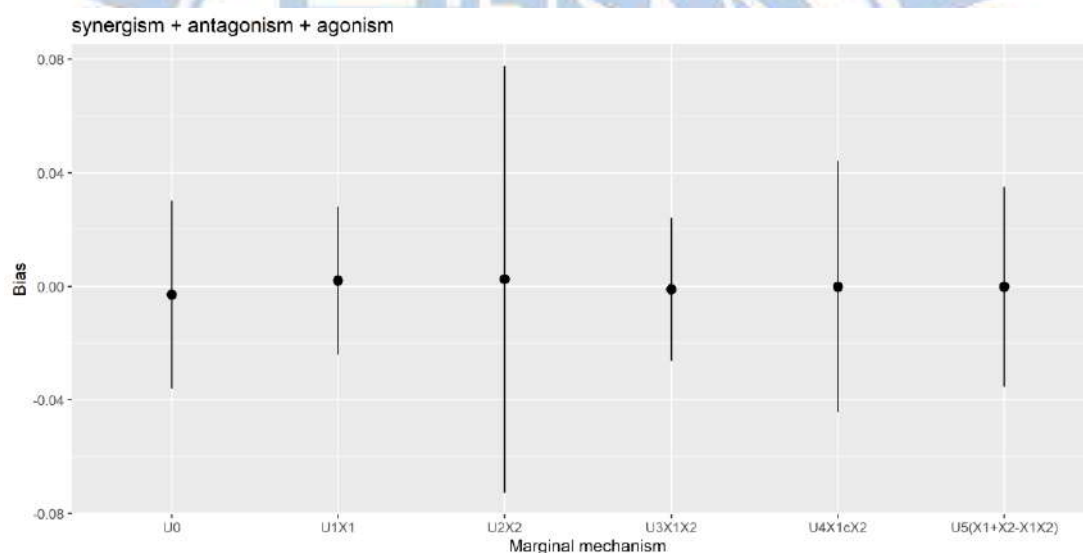


圖 3: 拮抗、促效、協同三種機制交互作用同時存在情況下，模擬的偏差值 95%信賴區間。

4.1.2. 考慮交互作用只有拮抗與協同兩種機制的存在

也就是結果Y會經由兩個因子的交互作用而發生。而這兩種交互作用機制，一種交互作用機制是會經由因子 X_1 和因子 X_2 的作用過程中，因子 X_1 會抑制因子

X_2 的效果。另一種是因子 X_1 的效果，會由於加上因子 X_2 的效果而產生加乘的效果。

在產生資料的步驟，把邊際機制的促效作用參數 $p_{U_{50},C=0}$ 、 $p_{U_{50},C=1}$ ，代理變數的參數 p_{Z_2} 設定為0。把估計方法用在產生的資料，可以得到以下五種機制導致結果發生的機率估計。不透過兩個風險因子使結果 Y 發生的機率為0.195 (0.162, 0.228)。經由因子 X_1 的邊際效果機制使結果 Y 發生的機率為0.085 (0.063, 0.107)。經由因子 X_2 的邊際效果機制使結果 Y 發生的機率為0.151 (0.082, 0.220)。經由兩個因子 X_1 、 X_2 的邊際協同機制使結果 Y 發生的機率為0.011 (-0.013, 0.036)。應由兩個因子 X_1 、 X_2 的邊際協同機制，因子 X_1 會抑制因子 X_2 ，然後再使結果 Y 發生的機率為0.032 (-0.012, 0.076)。

從表格6可以發現，兩個因子的主效用與拮抗、協同邊際機制交互作用，它們分別由機率模型參數所計算的理論值，與邊際成份原因模型估計值，兩者之間的差異(bias)都相當接近0。考慮模擬當中每一次重複估計的信賴區間，這五個機制的平均覆蓋率都接近95%。因此圖4的偏誤值信賴區間有95%的信心水準會包含0。這代表如果觀察到的資料如果在充分成份原因模型下有拮抗、協同兩種機制存在，可以由邊際成份原因模型估計這兩種邊際機制交互作用的大小。換句話說如果在細胞的尺度下有拮抗、協同兩種機制存在，則可以經由邊際成份原因模型估計在個體層級的兩種邊際機制的大小。

表格 6: 拮抗與協同二種機制交互作用同時存在的模擬結果。表格欄位由左而右分別是理論機率值、比例、估計機率值、偏差值、估計標準差、95%信賴區間上下界、模擬覆蓋率。

	Prob.	Prop.	est. prob.	bias	SE	LCI	UCI	coverage
mechanism U_0	0.194	41.0%	0.195	-0.001	0.017	-0.034	1.976	0.970
mechanism U_1X_1	0.085	18.0%	0.085	0.000	0.011	-0.022	1.972	0.970
mechanism U_2X_2	0.149	31.4%	0.151	-0.002	0.035	-0.071	1.993	0.960
synergism $U_3X_1X_2$	0.011	2.4%	0.011	0.000	0.012	-0.024	1.972	0.955
antagonism $U_{40}Z_1X_1^cX_2$	0.033	6.9%	0.032	0.001	0.022	-0.043	1.983	0.945

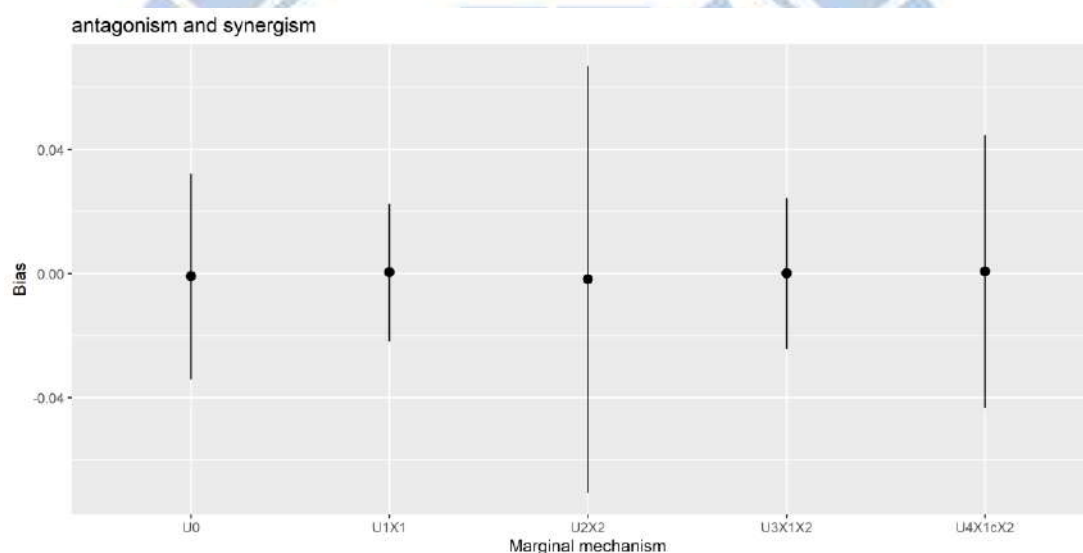


圖 4: 拮抗與協同二種機制交互作用同時存在情況下，模擬的偏差值 95%信賴區間。

4.1.3. 假設只有協同與促效交互作用機制

結果 Y 的發生，會經由因子 X_1 與因子 X_2 對結果有加成效果的邊際協同機制，與另一種會經由因子 X_1 與因子 X_2 兩者對結果 Y 的競爭的邊際促效機制。因此在產生資料的步驟，把邊際機制的拮抗作用參數 $p_{U_{40},C=0}$ 、 $p_{U_{40},C=1}$ ，代理變數的參數 p_{Z_1} 設定為 0。把估計的方法用在產生的資料上，可以知道不經由兩個因子使結果

的邊際機制，機率為 0.195 (0.162, 0.228)。只經由因子 X_1 使結果 Y 發生的邊際機制，它的機率為 0.086 (0.059, 0.112)。只經由因子 X_2 使結果 Y 發生的邊際機制，它的機率為 0.150 (0.109, 0.192)。由兩個風險因子 X_1 與 X_2 的邊際協同機制使結果發生的機率為 0.011 (-0.007, 0.028)。由兩個風險因子 X_1 與 X_2 經由對結果 Y 的競爭，在邊際促效機制下的機率為 0.043 (0.008, 0.078)。

從表格 7 可以發現，兩個因子的主效用與協同、促效邊際機制交互作用，它們分別由機率模型參數所計算的理論值，與邊際成份原因模型估計值，兩者之間的差異(bias)都相當接近 0。考慮模擬當中每一次重複估計的信賴區間，這五個機制的平均覆蓋率都接近 95%。因此圖 5 的偏誤值信賴區間有 95%的信心水準會包含 0。這代表如果觀察到的資料如果在充分成份原因模型下有協同、促效兩種機制存在，可以由邊際成份原因模型估計這兩種邊際機制交互作用的大小。換句話說如果在細胞的尺度下有協同、促效兩種機制存在，則可以經由邊際成份原因模型估計在個體層級的兩種邊際機制的大小。

表格 7: 協同與促效二種機制交互作用同時存在的模擬結果。表格欄位由左而右分別是理論機率值、比例、估計機率值、偏差值、估計標準差、95%信賴區間上下界、模擬覆蓋率。

	Prob.	Prop.	est. prob.	bias	SE	LCI	UCI	coverage
mechanism U_0	0.194	40.1%	0.195	-0.001	0.017	-0.034	1.976	0.970
mechanism U_1X_1	0.085	17.6%	0.086	0.000	0.013	-0.027	1.973	0.965
mechanism U_2X_2	0.149	30.7%	0.150	-0.001	0.021	-0.043	1.980	0.960
synergism $U_3X_1X_2$	0.011	2.3%	0.011	0.001	0.009	-0.017	1.969	0.945
agonism $U_{50}Z_2(X_1 + X_2 - X_1X_2)$	0.043	8.9%	0.043	0.000	0.018	-0.035	1.978	0.960

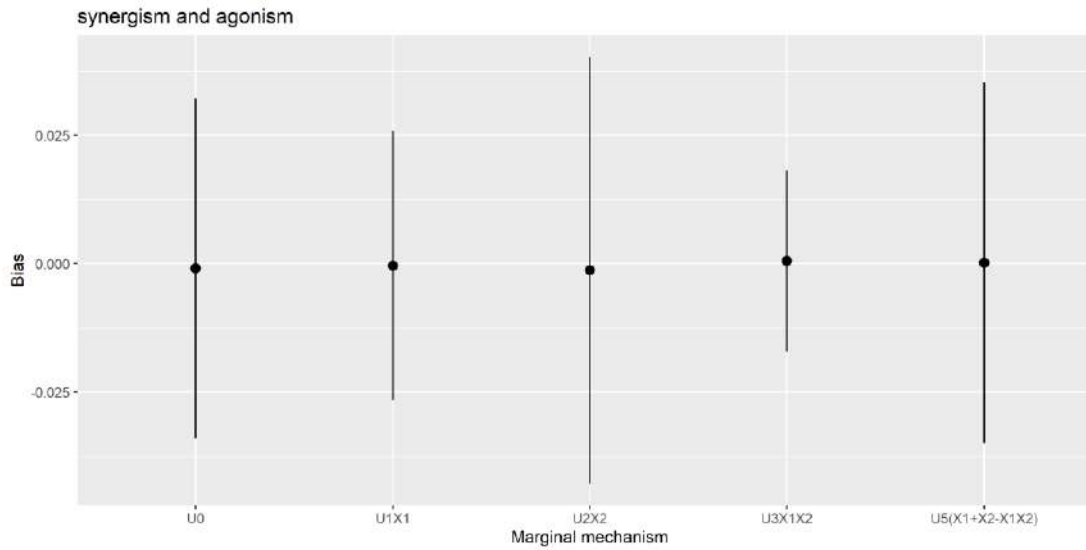


圖 5：協同與促效二種機制交互作用同時存在情況下，模擬的偏差值 95% 信賴區間。

4.1.4. 假設只有拮抗與促效

機率參數 $p_{U_3,C=0}$ 、 $p_{U_3,C=1}$ 、設定為 0。模擬的結果顯示，不經由任何風險因子造成結果 Y 發生的機率有 0.193(0.161, 0.226)。由單一風險因子 X_1 的邊際機制造成結果發生的機率有 0.086(0.011, 0.064)。由單一風險因子 X_2 的邊際機制造成結果發生的機率有 0.147 (0.107, 0.188)。由兩個風險因子 X_1 、 X_2 ，經由代理變數 Z_1 辨識的邊際拮抗機制，這個拮抗機制是因子 X_1 抑制 X_2 的作用，其造成結果發生的機率有 0.034(0.003, 0.065)。由兩個風險因子 X_1 、 X_2 ，經由代理變數 Z_2 辨識的邊際促效機制，這個機制是因子 X_1 與因子 X_2 競爭結果 Y 的發生，其造成結果發生的機率有 0.044(0.017, 0.010)。

從表格 8 可以發現，兩個因子的主效用與拮抗、促效邊際機制交互作用，它們分別由機率模型參數所計算的理論值，與邊際成份原因模型估計值，兩者之間

的差異(bias)都相當接近 0。考慮模擬當中每一次重複估計的信賴區間，這五個機制的平均覆蓋率都接近 95%。因此圖 6 的偏誤值信賴區間有 95%的信心水準會包含 0。這代表如果觀察到的資料如果在充分成份原因模型下有拮抗、促效兩種機制存在，可以由邊際成份原因模型估計這兩種邊際機制交互作用的大小。換句話說如果在細胞的尺度下有拮抗、促效兩種機制存在，則可以經由邊際成份原因模型估計在個體層級的兩種邊際機制的大小。

表格 8: 拮抗與促效二種機制交互作用同時存在的模擬結果。表格欄位由左而右分別是理論機率值、比例、估計機率值、偏差值、估計標準差、95%信賴區間上下界、模擬覆蓋率。

	Prob.	Prop.	est. prob.	bias	SE	LCI	UCI	coverage
mechanism U_0	0.194	38.5%	0.193	0.001	0.016	-0.031	1.977	0.945
mechanism U_1X_1	0.085	16.9%	0.086	0.000	0.011	-0.022	1.970	0.940
mechanism U_2X_2	0.149	29.5%	0.147	0.002	0.021	-0.039	1.982	0.950
antagonism $U_{40}Z_1X_1^cX_2$	0.033	6.5%	0.034	-0.001	0.016	-0.032	1.975	0.960
agonism $U_{50}Z_2(X_1 + X_2 - X_1X_2)$	0.043	8.5%	0.044	-0.001	0.017	-0.035	1.976	0.950

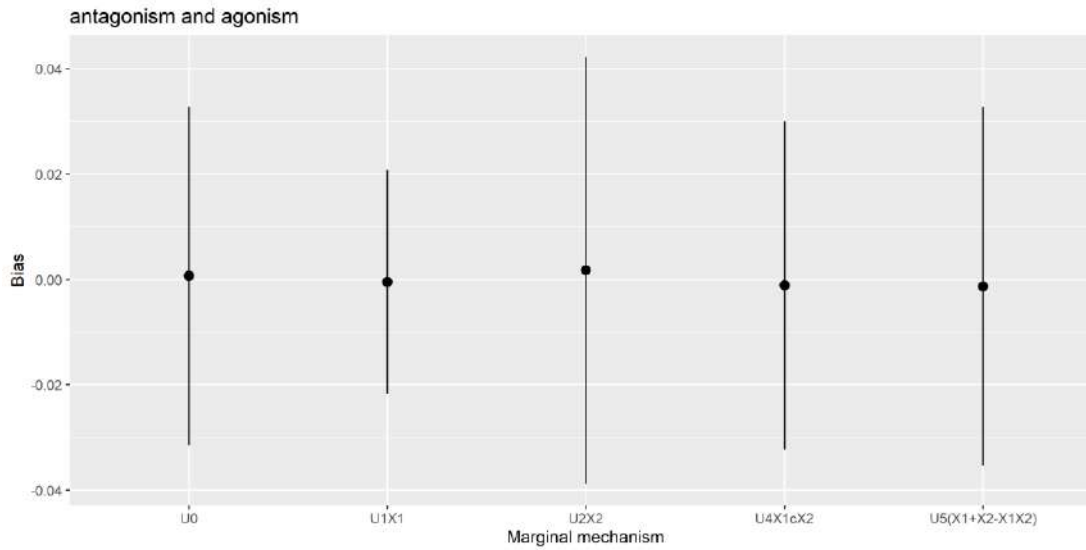


圖 6：拮抗與促效二種機制交互作用同時存在情況下，模擬的偏差值 95%信賴區間。

4.1.5. 假設只有拮抗

機率參數 $p_{U_3,C=0}$ 、 $p_{U_3,C=1}$ 、 $p_{U_{50},C=0}$ 、 $p_{U_{50},C=1}$ 設定為 0。模擬的結果顯示，不經由任何風險因子造成結果 Y 發生的機率有 0.193(0.160, 0.226)。由單一風險因子 X_1 的邊際機制造成結果發生的機率有 0.086(0.063, 0.108)。由單一風險因子 X_2 的邊際機制造成結果發生的機率有 0.150(0.101, 0.199)。由兩個風險因子 X_1 、 X_2 ，經由代理變數 Z_1 辨識的邊際拮抗機制，這個拮抗機制是因子 X_1 抑制 X_2 的作用，其造成結果發生的機率有 0.038(-0.003, 0.078)。

從表 9 可以發現，兩個因子的主效用與邊際拮抗機制交互作用，它們分別由機率模型參數所計算的理論值，與邊際成份原因模型估計值，兩者之間的差異(bias)都相當接近 0。考慮模擬當中每一次重複估計的信賴區間，這四個機制的平均覆蓋率都接近 95%。因此圖 7 的偏誤值信賴區間有 95%的信心水準會包含 0。

這代表如果觀察到的資料如果在充分成份原因模型下有拮抗機制存在，可以由邊際成份原因模型估計這邊際拮抗機制交互作用的大小。換句話說如果在細胞的尺度下有拮抗機制存在，則可以經由邊際成份原因模型估計在個體層級的邊際拮抗機制交互作用的大小。

表格 9：只有拮抗機制交互作用的模擬結果。表格欄位由左而右分別是理論機率值、比例、估計機率值、偏差值、估計標準差、95%信賴區間上下界、模擬覆蓋率。

	Prob.	Prop.	est. prob.	bias	SE	LCI	UCI	coverage
mechanism U_0	0.194	41.7%	0.193	0.001	0.017	-0.031	1.978	0.945
mechanism U_1X_1	0.085	18.3%	0.086	0.000	0.011	-0.022	1.971	0.970
mechanism U_2X_2	0.149	32.0%	0.150	-0.001	0.025	-0.051	1.984	0.970
antagonism $U_{40}Z_1X_1^cX_2$	0.036	7.8%	0.038	-0.001	0.021	-0.041	1.979	0.950

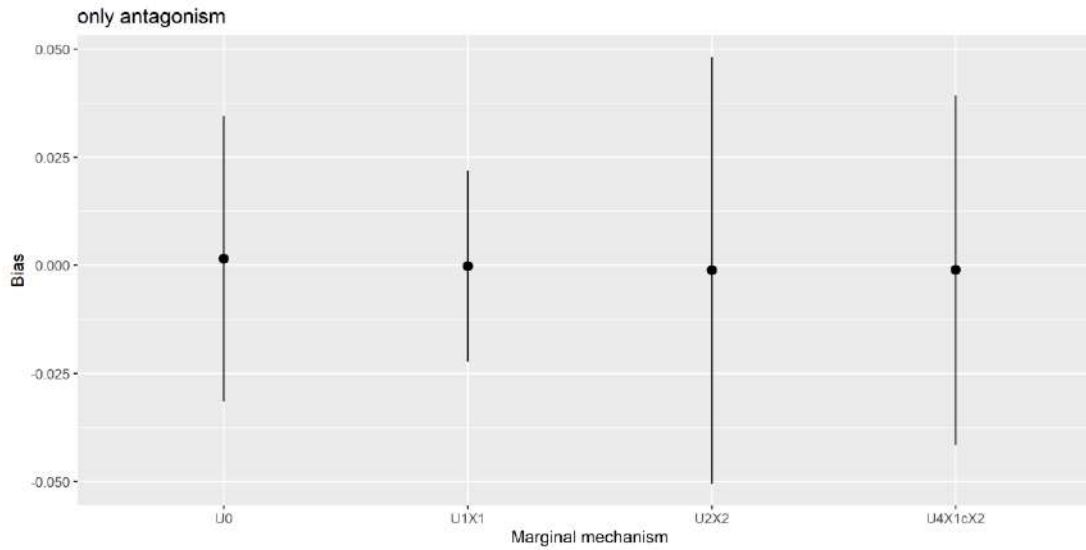


圖 7：只有拮抗機制交互作用情況下，模擬的偏差值 95%信賴區間。

4.1.6. 假設只有促效

機率參數 $p_{U_3,C=0}$ 、 $p_{U_3,C=1}$ 、 $p_{U_{40},C=0}$ 、 $p_{U_{40},C=1}$ 設定為 0。模擬的結果顯示，不經由任何風險因子造成結果Y發生的機率有 0.196(0.163, 0.229)。由單一風險因子 X_1 的邊際機制造成結果發生的機率有 0.084(0.067, 0.101)。由單一風險因子 X_2 的邊際機制造成結果發生的機率有 0.148(0.100, 0.196)。由兩個風險因子 X_1 、 X_2 ，經由代理變數 Z_2 辨識的邊際促效機制，這個機制是因子 X_1 與因子 X_2 競爭結果Y的發生，其造成結果發生的機率有 0.043(-0.023, 0.109)。

從表格 10 可以發現，兩個因子的主效用與邊際促效機制交互作用，它們分別由機率模型參數所計算的理論值，與邊際成份原因模型估計值，兩者之間的差異(bias)都相當接近 0。考慮模擬當中每一次重複估計的信賴區間，這四個機制的平均覆蓋率都接近 95%。因此圖 8 的偏誤值信賴區間有 95%的信心水準會包含

0。這代表如果觀察到的資料如果在充分成份原因模型下有促效機制存在，可以由邊際成份原因模型估計這邊際促效機制交互作用的大小。換句話說如果在細胞的尺度下有促效機制存在，則可以經由邊際成份原因模型估計在個體層級的邊際促效機制交互作用的大小。

表格 10：只有促效機制交互作用的模擬結果。表格欄位由左而右分別是理論機率值、比例、估計機率值、偏差值、估計標準差、95%信賴區間上下界、模擬覆蓋率。

	Prob.	Prop.	est. prob.	bias	SE	LCI	UCI	coverage
mechanism U_0	0.194	41.2%	0.196	-0.002	0.017	-0.035	1.975	0.955
mechanism U_1X_1	0.085	18.1%	0.084	0.001	0.009	-0.016	1.970	0.945
mechanism U_2X_2	0.149	31.6%	0.148	0.001	0.024	-0.047	1.985	0.950
agonism $U_{50}Z_2(X_1 + X_2 - X_1X_2)$	0.043	9.1%	0.043	0.000	0.034	-0.066	1.993	0.950

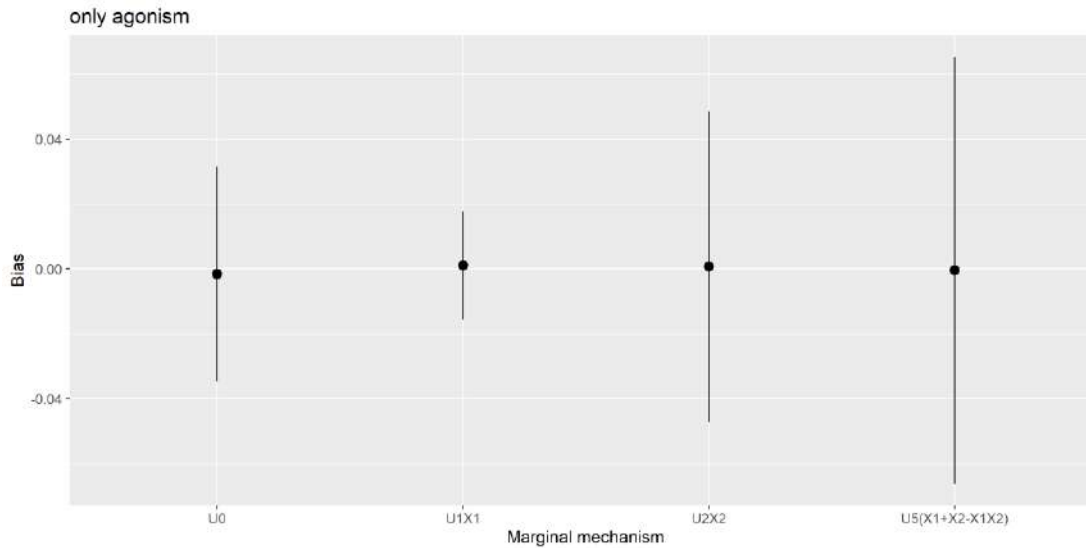


圖 8：只有促效機制交互作用情況下，模擬的偏差值 95%信賴區間。

4.1.7. 假設只有協同

機率參數 $p_{U_{40},C=0}$ 、 $p_{U_{40},C=1}$ 、 $p_{U_{50},C=0}$ 、 $p_{U_{50},C=1}$ 設定為 0。模擬的結果顯示，

不經由任何風險因子造成結果 Y 發生的機率 0.193(0.160, 0.226)。由單一風險因子 X_1 的邊際機制造成結果發生的機率 0.086(0.063, 0.108)。由單一風險因子 X_2 的邊際機制造成結果發生的機率有 0.151(0.122, 0.180)。由兩個風險因子 X_1 、 X_2 的邊際協同機制造成結果發生的機率有 0.011(-0.005, 0.027)。

從表格 11 可以發現，兩個因子的主效用與邊際協同機制交互作用，它們分別由機率模型參數所計算的理論值，與邊際成份原因模型估計值，兩者之間的差異(bias)都相當接近 0。考慮模擬當中每一次重複估計的信賴區間，這四個機制的平均覆蓋率都接近 95%。因此圖 8 的偏誤值信賴區間有 95%的信心水準會包含 0。這代表如果觀察到的資料如果在充分成份原因模型下有協同機制存在，可以由邊際成份原因模型估計這邊際協同機制交互作用的大小。換句話說如果在細胞的尺度下有協同機制存在，則可以經由邊際成份原因模型估計在個體層級的邊際協同機制交互作用的大小。

表格 11：只有協同機制交互作用的模擬結果。表格欄位由左而右分別是理論機率值、比例、估計機率值、偏差值、估計標準差、95%信賴區間上下界、模擬覆蓋率。

	Prob.	Prop.	est. prob.	bias	SE	LCI	UCI	coverage
mechanism U_0	0.194	44.1%	0.193	0.001	0.017	-0.031	1.978	0.945
mechanism U_1X_1	0.085	19.4%	0.086	0.000	0.011	-0.022	1.971	0.970
mechanism U_2X_2	0.149	33.8%	0.151	-0.002	0.015	-0.031	1.972	0.945
synergism $U_3X_1X_2$	0.011	2.6%	0.011	0.000	0.008	-0.016	1.968	0.960

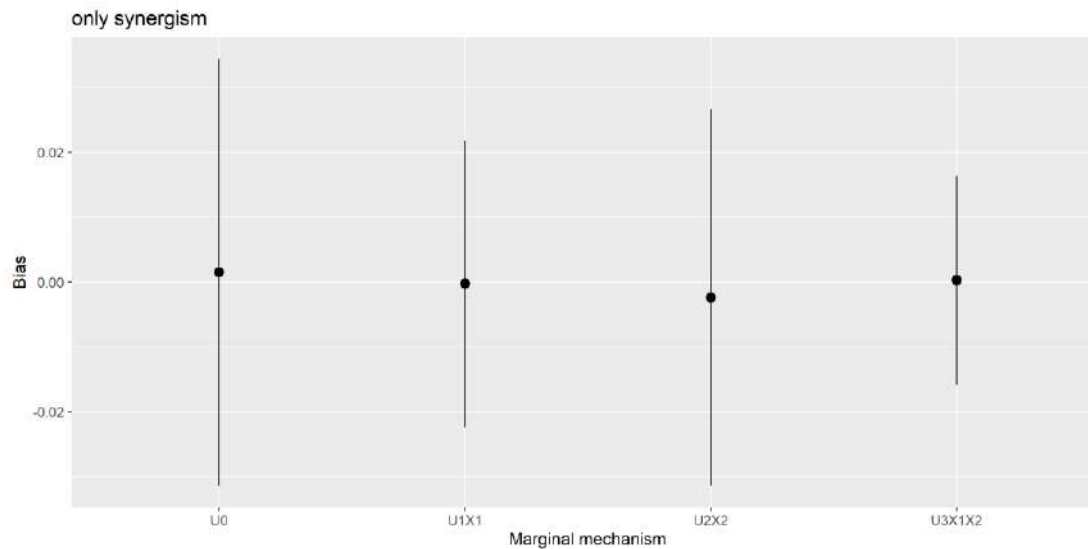


圖 9：只有協同機制同時存在情況下，模擬的偏差值 95%信賴區間。

從每次由應用估計方法在理論模型產生的資料所得到的信賴區間，計算平均覆蓋率接近 0.95，這說明在這個參數的設定下，估計方法可以適當的估計各邊際機制的對 Y_U 發生的理論機率。

4.2. 資料分析

本研究分析的資料為台灣肝癌資料，它是 1991 與 1992 的世代研究，追蹤到 2004 年。資料從 7 個鄉鎮蒐集，樣本的年齡分佈從 30 到 65 歲。這份自資料共有 23820 筆的紀錄，有 8 個欄位。這 8 個欄位代表意義與名稱分別是個體是否罹患肝癌(hcc.case)、性別(GENDER)、肝臟發炎指標的血清轉胺酶 1 (alt1)、肝臟發炎指標的血清轉胺酶 2 (alt2)、個體是否酗酒(alcohol1)、個體的 B 型肝炎病毒 DNA 量 (HBV_DNA1)、個體是否檢驗出 C 型肝炎病毒抗體(ANTIHCV)、個體是否檢驗出 B 型肝炎病毒表面抗原(HBV_sAg)。欄位 hcc.case 是二元變數，1 代表個體

罹患肝癌，反之為 0。兩個欄位 alt 代表肝臟是否發炎狀態，1 代表是，反之為 0。

欄位 alcohol1 代表個體是否酗酒，1 代表是，反之為 0。欄位 HBVsAg 代表個體是否感染 B 肝病毒，1 代表是，反之為 0。欄位 ANTIHCV 代表個體是否感染 C 肝病毒，1 代表是，反之為 0。

在這 23820 個樣本中，有 2.153 % (1.969 %, 2.338 %) 的比例罹患 HCC 肝癌，17.443 % (16.961 %, 17.925 %) 的比例感染 B 型肝炎病毒，5.512 % (5.222 %, 5.801 %) 的比例感染 C 型肝炎病毒，30.188 % (29.605 %, 30.771 %) 的比例有肝臟發炎，10.604 % (10.213 %, 10.995 %) 的比例有酗酒習慣。在罹患肝癌的樣本當中，60.233 % (55.998 %, 64.469 %) 的比例感染 B 型肝炎病毒，26.510 % (22.691 %, 30.330 %) 的比例感染 C 型肝炎病毒，66.276 % (62.185 %, 70.367 %) 的比例有肝臟發炎，18.713 % (15.338 %, 22.088 %) 的比例有酗酒習慣。

這裡我們邊際成份原因模型用在此筆資料上，建立模型來分析 B 型肝炎與 C 型肝炎病毒導致肝癌的機制。對於模型變數的部分，所有的變數都是二元變數。 X_1 代表個體是否有 C 型肝炎病毒， $X_1 = 1$ 表示個體感染了 C 肝病毒， $X_1 = 0$ 代表沒有。 X_2 代表個體是否有 B 型肝炎病毒， $X_2 = 1$ 表示個體感染了 B 肝病毒， $X_2 = 0$ 代表沒有。 Y 代表個體是否有肝癌 HCC， $Y = 1$ 代表個體罹患肝癌， $Y = 0$ 代表沒有。代理變數 Z_1 代表個體的 B 肝病毒數量是否大於門檻值 (HBV DNA = 10000000)， $Z_1 = 1$ 代表大於門檻值， $Z_1 = 0$ 代表沒有。代理變數 Z_2 代表個體是否有飲酒習慣 [9, 32]， $Z_2 = 1$ 代表個體有飲酒習慣， $Z_2 = 0$ 則代表沒有。變數 C

代表性別，它是因果圖的干擾因子。關於代理變數的選擇有兩個原因。選用 B 肝病毒數量是否大於門檻值來當作代理變數 Z_1 ，原因是考慮了 C 肝病毒抑制 B 肝病毒的拮抗機制作用，需要有一定的病毒量來使肝癌發生，所以當病毒量大於門檻值，可以相對保守的說明拮抗作用導致肝癌。然後分別對兩種情況下作機制的估計。從結果來看，造成肝癌的可能機制當中，除了其它原因、只經由 C 或 B 肝病毒的機制這三種，經由拮抗機制造成肝癌也是顯著的。以下說明兩種模型情況的資料分析結果。

4.2.1. 只考慮協同與促效兩種機制

不考慮拮抗機制的情況下，從表格 13 的結果來看各機制所佔的比例與顯著程度。罹患肝癌的個體當中並且顯著的機制有三個。約有 28%(風險 0.0055，信賴區間(0.0044, 0.0066)的比例是經由其它機制。約有 23%(風險 0.0046，信賴區間(0.0036, 0.0056)的比例是只經由 C 型肝炎病毒的機制。約有 43%(風險 0.0083，信賴區間(0.007, 0.0096)的比例是只經由 B 型肝炎病毒的機制。不顯著但是如不可忽略的機制有一個，約有 3.7%(風險 0.0007，信賴區間(-0.0001, 0.0016)的比例是經由 C 肝病毒與 B 肝病毒的促效機制。不顯著的機制有一個，約有 1%(風險 0.0002，信賴區間(-0.0003, 0.0007)的比例是經由 C 肝病毒與 B 肝病毒的協同機制。

表格 12：考慮 HCV 與 HBV 導致 HCC 經由促效、協同兩種機制存在的情況下，由邊際成份原因模型所估計的各機制大小。欄位由左而右分別是機率、比例、

估計誤差、95%信賴區間上下界、P 值。

	est. prob.	est. prop.	SE	LCI	UCI	P-value
other mechanism	0.0055	28.4%	0.0006	0.0044	0.0066	<0.001
HCV only	0.0046	23.8%	0.0005	0.0036	0.0056	<0.001
HBV only	0.0083	43.0%	0.0007	0.0070	0.0096	<0.001
HCV HBV synergism	0.0002	1.1%	0.0002	-0.0003	0.0007	0.3886
HCV HBV antagonism	-	-	-	-	-	-
HCV HBV agonism	0.0007	3.7%	0.0004	-0.0001	0.0016	0.0957

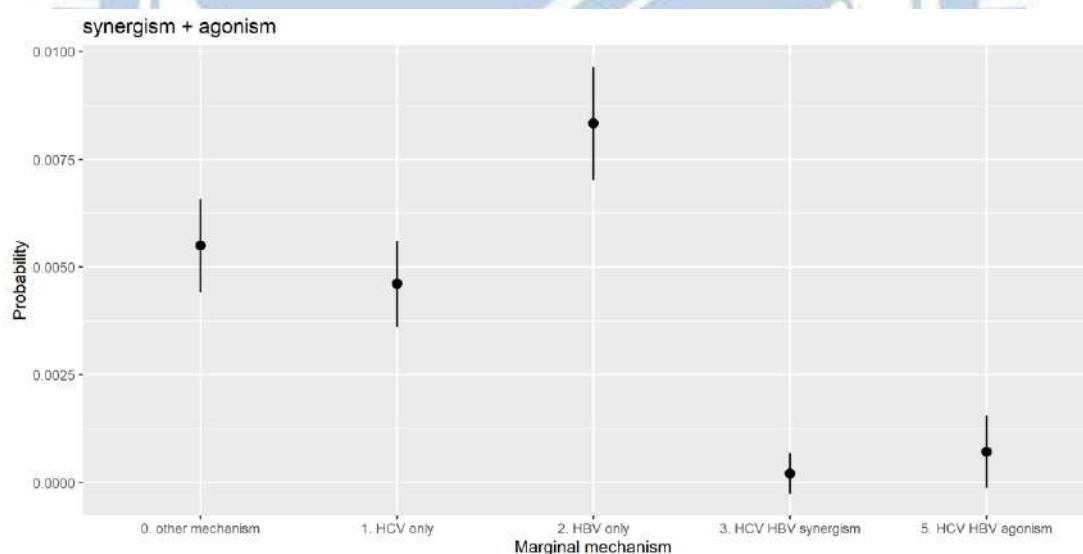


圖 10：考慮 HCV 與 HBV 導致 HCC 經由促效、協同兩種機制作用存在情況下，由邊際成份原因模型估計的 95%信賴區間。

4.2.2. 拮抗、協同與促效三種機制

從表格 12 的結果來看各機制所佔的比例與顯著程度。在邊際充分原因模型

的架構下，在罹患肝癌的個體當中並且顯著的機制有四個。約有 31%(風險 0.0055，信賴區間(0.0044, 0.0066)的比例是經由其它機制。約有 26%(風險 0.0046，信賴區間(0.0036, 0.0056)的比例是只經由 C 型肝炎病毒的機制。約有 35%(風險 0.0062，信賴區間(0.005, 0.0074)的比例是只經由 B 型肝炎病毒的機制。約有 1.8%(風險 0.0003，信賴區間(0.0002, 0.0004)的比例是經由 C 肝病毒與 B 肝病毒的拮抗機制。雖然不顯著但是如不可忽略的機制有兩個。約有 1.7%(風險 0.0003，信賴區間(-0.0002, 0.0008)的比例是經由 C 肝病毒與 B 肝病毒的協同機制。約有 3.8%(風險 0.0007，信賴區間(-0.0002, 0.0015)的比例是經由 C 肝病毒與 B 肝病毒的促效機制。

表格 13：考慮 HCV 與 HBV 導致 HCC 經由拮抗、促效、協同三種機制存在的情況下，由邊際成份原因模型所估計的各機制大小。欄位由左而右分別是機率、比例、估計誤差、95%信賴區間上下界、P 值。

	est. prob.	est. prop.	SE	LCI	UCI	P-value
other mechanism	0.0055	31.2%	0.0006	0.0044	0.0066	<0.001
HCV only	0.0046	26.2%	0.0005	0.0036	0.0056	<0.001
HBV only	0.0062	35.1%	0.0006	0.0050	0.0074	<0.001
HCV HBV synergism	0.0003	1.8%	0.0002	-0.0002	0.0008	0.1901
HCV HBV antagonism	0.0003	1.9%	0.0001	0.0002	0.0004	<0.001
HCV HBV agonism	0.0007	3.8%	0.0004	-0.0002	0.0015	0.1202

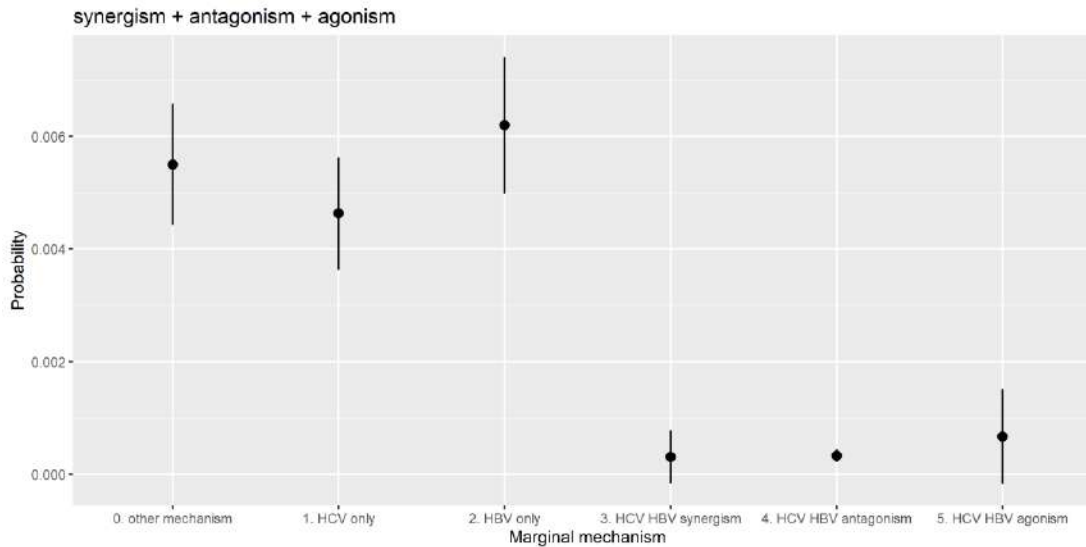


圖 11：考慮 HCV 與 HBV 導致 HCC 經由拮抗、促效、協同三種機制作用存在情況下，由邊際成份原因模型估計的 95%信賴區間。

從兩種模型假設結果的資料分和表格 14 比較可以發現，如果不考慮拮抗機制存在，經由拮抗機制對肝癌作用的機制大小，會分別對透過 B 肝病毒造成肝癌的機制的估計放大，也會對透過 C 肝病毒與 B 肝病毒兩者經由協同機制，造成肝癌的協同機制的估計量減少，使協同作用被低估。

如果模型考慮拮抗機制存在，在這筆資料分析中可以發現 C 肝與 B 肝兩種病毒造成肝癌的交互作用，協同與促效兩種機制是不顯著的，然而 C 肝與 B 肝兩種病毒經由拮抗機制導致肝癌發生雖然估計量只有 0.0003，它是顯著的。雖然協同作用與促效作用的結果不顯著，但是從估計量來看，它們大於等於拮抗機制的估計量，而且對應的 p 值也並非很大。然而拮抗作用的估計量雖然相當微量，但是它的檢定結果顯示相當顯著，代表 C 肝與 B 肝這兩種病毒經由拮抗作用導致肝癌發生的機制是重要的。

表格 14：比較有拮抗機制(欄位左)與沒有拮抗機制(欄位右)的邊際機制的背景成份理論值大小。可以發現在沒有拮抗機制的情況下(欄位右)，拮抗機制的期望值會在因子 X2 的效果與協同機制效果這兩者分別出現。

	Antagonistic & agonistic & synergistic mechanisms	Agonistic & synergistic mechanisms
Other	$E[U_0 C]$	$E[U_0 C]$
X₁ only	$E[U_1 C]E[X_1 C]$	$E[U_1 C]E[X_1 C]$
X₂ only	$E[U_2 C]E[X_2 C]$	$(E[U_2 C] + E[U_{40} C]E[Z_1])E[X_2 C]$
X₁ and X₂ Synergism	$E[U_3 C]E[X_1X_2 C]$	$(E[U_3 C] - E[U_{40} C]E[Z_1])E[X_1X_2 C]$
X₁ and X₂ Antagonism	$E[U_{40} C]E[Z_1]E[(1 - X_1)X_2 C]$	None
X₁ and X₂ Agonsim	$E[U_{50} C]E[Z_2]E[X_1 + X_2 - X_1X_2 C]$	$E[U_{50} C]E[Z_2]E[X_1 + X_2 - X_1X_2 C]$

5. 討論

本研究延伸了學者 Lin, et al. [9]的邊際充分原因成份模型估計機制的方法，在充分成份原因之下，釋放因子成份對結果的預防效用假設，增加在邊際成份原因模型可估計的機制。假設拮抗、促效、協同三種機制交互作用分別只在模型出現一次，也就是不會有兩種以上的拮抗機制，可以經由本研究提出的邊際充分原因模型來同時估計與檢定這三種機制交互作用。

觀察背景成份向量與反應模式的對應，可以發現邊際拮抗機制與邊際協同

機制的反應模式剛好對應到一組背景成份向量。此向量的元素從充分成份原因模型的角度來看，分別各自單獨對應拮抗機制與協同機制的背景成份因(表格 2)。因此估計的邊際拮抗或是協同機制如果顯著，可以說明在充分成份原因模型下的拮抗或是協同機制存在。而估計的量也可以分別代表拮抗或協同機制大小的下界。

然而考慮充分成份原因的背景成份向量，在兩個風險因子的情況下，向量有 $2^3 + 1 = 9$ 個元素，這些二元變數的元素產生 $2^9 = 512$ 種情況，對應到兩個風險因子 $2^2 = 4$ 個元素組成的反事實結果向量，因此可以有 $2^4 = 16$ 種反應模式。如果要估計這 16 種反應模式對應的機制。在兩個風險因子，兩個代理變數的線性飽和模型之下，至多只能估計 $2^3 + 1 = 9$ 機制。因此要估計所有機制，需要加入兩個以上的代理變數。在三個風險因子的情況下，向量有 $3^3 + 1 = 28$ 個元素，這些二元變數的元素產生 2^{28} 種情況，對應到三個風險因子 $2^3 = 8$ 個元素組成的反事實結果向量，因此可以有 $2^8 = 256$ 種反應模式。同樣的概念，在三個風險因子，三個代理變數的線性飽和模型之下，至多只能估計 $3^3 + 1 = 28$ 機制。因此要估計所有機制，這時需要的代理變數至少需要三個以上。隨著考慮的風險因子增加，反應模式的種類也會大幅增加。假設有 k 個風險因子，這時的背景成份向量有 $k^3 + 1$ 個元素，這些二元變數的元素產生 2^{k^3+1} 種情況，對應到 k 個風險因子 2^k 個元素組成的反事實結果向量，因此可以有 2^{2^k} 種反應模式。隨著風險因子的增加，要估計的邊際機制會大幅度的增加，需要的代理變數個數與機制增加的量不成比例。

引此在使用邊際成份原因模型前，必須先確定估計目標，經由非有向因果關係圖事先確認變數之間的因果關係，盡可能把風險因子降到兩個，才能在少量的代理變數之下估計出目標機制。

另一限制是代理變數的選擇必須在資料蒐集前一併考慮或是可能的話在實驗過程一併測量。由於代理變數在同時估計多個邊際機制交互作用的必須成份，預先考慮需要的代理變數，將可以在資料蒐集之後估計較多的候選機制。

由於邊際成份原因模型的結果，是經由充分原因模型各機制是否發生而來量化，經過反事實結果與因果推論的一致性和可交換性假設，從反事實結果的期望值對應到統計模型的期望值。然而統計模型的期望值由線性模型來建模，使得兩邊對應的函數值域不一樣。自變數的係數變化範圍可以是在 $(0, 1)$ 區間之外，但是二元變數的反事實結果的期望值範圍只在 $(0, 1)$ 。因此可以藉由學者 Lin, et al. [10]的方法，分析問題前於蒐集資料或實驗的階段把時間的標記列入考慮，如此一來可以把原本估計的對象改為風險函數，改由在存活函數的尺度之下來估計與檢定各機制的大小。如此便可以解決等式兩邊範圍不相等的問題。

附錄：資料模擬的生成細節

干擾因子的機率分佈為 $C \sim \text{Bernoulli}(p_C)$ ，其中 p_C 代表機率 $\Pr(C = 1)$ 。

風險因子 X_1 的機率分佈為 $X_1 \sim \text{Bernoulli}(p_{X_1})$ ，其中 p_{X_1} 代表 $\Pr(X_1 = 1)$ 的機率。

風險因子 X_2 的機率分佈為 $X_2 \sim \text{Bernoulli}(p_{X_2})$ ，其中 p_{X_2} 代表 $\Pr(X_2 = 1)$ 的機率。

代理變數 Z_1 的機率分佈為 $Z_1 \sim \text{Bernoulli}(p_{Z_1})$ ，其中 p_{Z_1} 代表 $\Pr(Z_1 = 1)$ 的機率。

代理變數 Z_2 的機率分佈為 $Z_2 \sim \text{Bernoulli}(p_{Z_2})$ ，其中 p_{Z_2} 代表 $\Pr(Z_2 = 1)$ 的機率。

由式 3.5 可以知道結果變數 Y_U 需要表示成 S_0 、 S_1 、 S_2 、 S_3 、 S_4 、 S_5 六個邊際成份原因加總。由式 3.6 可以知道要模擬這六個邊際成份原因，必須模擬它們各自對應的背景成份原因 U_0 、 U_1 、 U_2 、 U_3 、 U_{40} 、 U_{50} 。如果把背景條件 U 以向量的方式呈現 $(U_0, U_1, U_2, U_3, U_{40}, U_{50}, U_{-1})$ ，其中每個元素對應各機制，因此機率分配 $U \sim \text{multinomial}(n, p_{U_0}, p_{U_1}, p_{U_2}, p_{U_3}, p_{U_{40}}, p_{U_{50}}, p_{U_{-1}})$ 可以改以向量呈現 $(U_0, U_1, U_2, U_3, U_{40}, U_{50}, U_{-1}) \sim \text{multinomial}(n, p_{U_0}, p_{U_1}, p_{U_2}, p_{U_3}, p_{U_{40}}, p_{U_{50}}, p_{U_{-1}})$ 。其中 p_{U_0} 是 $\Pr(U_0 = 1)$ 的機率， p_{U_1} 是 $\Pr(U_1 = 1)$ 的機率， p_{U_2} 是 $\Pr(U_2 = 1)$ 的機率， p_{U_3} 是 $\Pr(U_3 = 1)$ 的機率， $p_{U_{40}}$ 是 $\Pr(U_{40} = 1)$ 的機率， $p_{U_{50}}$ 是 $\Pr(U_{50} = 1)$ 的機率， $p_{U_{-1}}$ 是 $\Pr(U_{-1} = 1)$ 的機率。

在不考慮干擾因子影響的情況下，依照以下步驟來模擬結果變數 Y 。第一步驟先給定資料個數 N ，然後分別設定風險因子的機率 p_{X_1} 、 p_{X_2} ，代理變數的機率

p_{z_1} 、 p_{z_2} ，背景成份原因的機率 p_{U_0} 、 p_{U_1} 、 p_{U_2} 、 p_{U_3} 、 $p_{U_{40}}$ 、 $p_{U_{50}}$ 、 $p_{U_{-1}}$ 。第二步驟去模擬每個參數對應的隨機變數，風險因子的值 x_1 與 x_2 、代理變數的值 z_1 與 z_2 ，背景成份原因的值 u_0 、 u_1 、 u_2 、 u_3 、 u_{40} 、 u_{50} 。第三步驟去計算由隨機變數模擬產生的各個邊際背景成份原因 $S_0 = s_0 = u_0$ 、 $S_1 = s_1 = u_1x_1$ 、 $S_2 = s_2 = u_2x_2$ 、 $S_3 = s_3 = u_3x_1x_2$ 、 $S_4 = s_4 = u_{40}z_1(1 - x_1)x_2$ 、 $S_5 = s_5 = u_{50}z_2(x_1 + x_2 - x_1x_2)$ 。第四步驟為計算 $Y_U = y_u$ 的實際值。參考式子 3.5，把各個模擬出的邊際成份原因加總可以得到 $y = s_0 + s_1 + s_2 + s_3 + s_4 + s_5$ 。第五步驟使用本研究方法計算各邊際成份原因的估計值，並且計算 Jackknife 變異數估計值[33, 34]。最後步驟重複上述一到五步驟 B 次數，計算個邊際成份原因的平均值、估計誤差、信賴區間以及覆蓋率。

在考慮干擾因子 C 影響的情況下，必須依照因子 C 的分層來模擬結果變數 Y 。模擬的過程與不考慮干擾因子的方法類似，差異在模擬變數 Y 的背景成份因 U 的時候，需要依照情境 $C = 0$ 與 $C = 1$ 各自產生數值。第一步驟先給定資料個數 N ，然後分別設定風險因子的機率 p_{x_1} 、 p_{x_2} ，代理變數的機率 p_{z_1} 、 p_{z_2} ，干擾因子在 $C = 0$ 分層的背景成份原因的機率 $p_{U_0,C=0}$ 、 $p_{U_1,C=0}$ 、 $p_{U_2,C=0}$ 、 $p_{U_3,C=0}$ 、 $p_{U_{40},C=0}$ 、 $p_{U_{50},C=0}$ 、 $p_{U_{-1},C=0}$ ，干擾因子在 $C = 1$ 分層的背景成份原因的機率 $p_{U_0,C=1}$ 、 $p_{U_1,C=1}$ 、 $p_{U_2,C=1}$ 、 $p_{U_3,C=1}$ 、 $p_{U_{40},C=1}$ 、 $p_{U_{50},C=1}$ 、 $p_{U_{-1},C=1}$ 。第二步驟去模擬每個參數對應的隨機變數，風險因子的值 x_1 與 x_2 、代理變數的值 z_1 與 z_2 ，干擾因子在 $C = 0$ 分層的背景成份原因的數值 $u_{0,C=0}$ 、 $u_{1,C=0}$ 、 $u_{2,C=0}$ 、 $u_{3,C=0}$ 、 $u_{40,C=0}$ 、 $u_{50,C=0}$ ，干擾因子在 $C =$

1分層的成份原因的數值 $u_{0,C=1}$ 、 $u_{1,C=1}$ 、 $u_{2,C=1}$ 、 $u_{3,C=1}$ 、 $u_{40,C=1}$ 、 $u_{50,C=1}$ 。第三步驟去計算由隨機變數模擬產生的各個邊際背景成份原因。干擾因子在 $C = 0$ 分層的 $S_{0,C=0} = s_{0,C=0} = u_{0,C=0}$ 、 $S_{1,C=0} = s_{1,C=0} = u_{1,C=0}x_1$ 、 $S_{2,C=0} = s_{2,C=0} = u_{2,C=0}x_2$ 、 $S_{3,C=0} = s_{3,C=0} = u_{3,C=0}x_1x_2$ 、 $S_{4,C=0} = s_{4,C=0} = u_{40,C=0}z_1(1-x_1)x_2$ 、 $S_{5,C=0} = s_{5,C=0} = u_{50,C=0}z_2(x_1+x_2-x_1x_2)$ 。干擾因子在 $C = 1$ 分層的 $S_{0,C=1} = s_{0,C=1} = u_{0,C=1}$ 、 $S_{1,C=1} = s_{1,C=1} = u_{1,C=1}x_1$ 、 $S_{2,C=1} = s_{2,C=1} = u_{2,C=1}x_2$ 、 $S_{3,C=1} = s_{3,C=1} = u_{3,C=1}x_1x_2$ 、 $S_{4,C=1} = s_{4,C=1} = u_{40,C=1}z_1(1-x_1)x_2$ 、 $S_{5,C=1} = s_{5,C=1} = u_{50,C=1}z_2(x_1+x_2-x_1x_2)$ 。第四步驟為計算 $Y = y$ 的實際值。參考式子 3.5 並且依照產生的干擾因子 $C = 0$ 與 $C = 1$ 分層，分別把模擬出的邊際成份原因加總可以得到 $y_{C=0} = s_{0,C=0} + s_{1,C=0} + s_{2,C=0} + s_{3,C=0} + s_{4,C=0} + s_{5,C=0}$ 與 $y_{C=1} = s_{0,C=1} + s_{1,C=1} + s_{2,C=1} + s_{3,C=1} + s_{4,C=1} + s_{5,C=1}$ 。第五步驟使用本研究方法計算各邊際成份原因的估計值，並且計算 Jackknife 的變異數估計值。最後步驟重複上述一到五步驟 B 次數，計算個邊際成份原因的平均值、估計誤差、信賴區間以及覆蓋率。

在不考慮干擾因子的情況之下，各邊際成份原因的理论機率為

$$\Pr(S_0 = 1) = p_{U_0}$$

$$\Pr(S_1 = 1) = p_{U_1}p_{X_1}$$

$$\Pr(S_2 = 1) = p_{U_2}p_{X_2}$$

$$\Pr(S_3 = 1) = p_{U_3}p_{X_1}p_{X_2}$$

$$\Pr(S_4 = 1) = p_{U_{40}} p_{Z_1} (1 - p_{X_1}) p_{X_2}$$

$$\Pr(S_5 = 1) = p_{U_{50}} p_{Z_2} (p_{X_1} + p_{X_2} - p_{X_1} p_{X_2})$$

考慮干擾因子分層效果，各邊際成份原因的理論機率，經由每個分層下去加權各機制的總合而得到。

$$\Pr(S_0 = 1) = (1 - p_C) \Pr(S_{0,C=0} = 1) + p_C \Pr(S_{0,C=1} = 1)$$

$$= (1 - p_C) p_{U_{0,C=0}} + p_C p_{U_{0,C=1}}$$

$$\Pr(S_1 = 1) = (1 - p_C) \Pr(S_{1,C=0} = 1) + p_C \Pr(S_{1,C=1} = 1)$$

$$= (1 - p_C) p_{U_{1,C=0}} p_{X_1} + p_C p_{U_{1,C=1}} p_{X_1}$$

$$\Pr(S_2 = 1) = (1 - p_C) \Pr(S_{2,C=0} = 1) + p_C \Pr(S_{2,C=1} = 1)$$

$$= (1 - p_C) p_{U_{2,C=0}} p_{X_2} + p_C p_{U_{2,C=1}} p_{X_2}$$

$$\Pr(S_3 = 1) = (1 - p_C) \Pr(S_3 = 1) + p_C \Pr(S_3 = 1)$$

$$= (1 - p_C) p_{U_{3,C=0}} p_{X_1} p_{X_2} + p_C p_{U_{3,C=1}} p_{X_1} p_{X_2}$$

$$\Pr(S_4 = 1) = (1 - p_C) \Pr(S_{4,C=0} = 1) + p_C \Pr(S_{4,C=1} = 1)$$

$$= (1 - p_C) p_{U_{40,C=0}} p_{Z_1} (1 - p_{X_1}) p_{X_2} + p_C p_{U_{40,C=1}} p_{Z_1} (1 - p_{X_1}) p_{X_2}$$

$$\Pr(S_5 = 1) = (1 - p_C) \Pr(S_{5,C=0} = 1) + p_C \Pr(S_{5,C=1} = 1) = (1 -$$

$$p_C) p_{U_{50,C=0}} p_{Z_2} (p_{X_1} + p_{X_2} - p_{X_1} p_{X_2}) + p_C p_{U_{50,C=1}} p_{Z_2} (p_{X_1} + p_{X_2} - p_{X_1} p_{X_2})。$$

知道參數值對應到機制的理論值，可以比較經由模擬的資料所估計的機制大

小，和理論值之間的差異如何。以下模擬分析為在邊際成份原因模型下，考慮拮抗、促效、協同三種機制的存在與否的可能組合，六種情況下的邊際成份原因估計值。

兩個風險因子的機率參數設定分別是 $p_{X_1} = 0.29$ 、 $p_{X_2} = 0.61$ 。

兩個代理變數的機率參數設定分別是 $p_{Z_1} = 0.9$ 、 $p_{Z_2} = 0.8$ 。

干擾因子的機率為 $p_C = 0.41$ 。

干擾因子在 $C = 0$ 分層的背景成份原因的機率設定為 $p_{U_0,C=0} = 0.19$ 、 $p_{U_1,C=0} = 0.29$ 、 $p_{U_2,C=0} = 0.24$ 、 $p_{U_3,C=0} = 0.06$ 、 $p_{U_{40},C=0} = 0.08$ 、 $p_{U_{50},C=0} = 0.07$ 、 $p_{U_{-1},C=0} = 0.07$ 。

干擾因子在 $C = 1$ 分層的背景成份原因的機率設定為 $p_{U_0,C=1} = 0.2$ 、 $p_{U_1,C=1} = 0.3$ 、 $p_{U_2,C=1} = 0.25$ 、 $p_{U_3,C=1} = 0.07$ 、 $p_{U_{40},C=1} = 0.09$ 、 $p_{U_{50},C=1} = 0.08$ 、 $p_{U_{-1},C=1} = 0.01$ 。

根據以上的參數設定，考慮干擾因子影響之下，可以計算邊際充分原因的各機制的理論機率。可以得到 $\Pr(S_0 = 1) = 0.194$ 、 $\Pr(S_1 = 1) = 0.085$ 、 $\Pr(S_2 = 1) = 0.149$ 、 $\Pr(S_3 = 1) = 0.011$ 、 $\Pr(S_4 = 1) = 0.033$ 、 $\Pr(S_5 = 1) = 0.043$ 。

參考文獻

1. VanderWeele, T.J., *Invited Commentary: The Continuing Need for the Sufficient Cause Model Today*. American Journal of Epidemiology, 2017. **185**(11): p. 1041-1043.
2. Rothman, K.J., *CAUSES*. American Journal of Epidemiology, 1976. **104**(6): p. 587-592.
3. VanderWeele, T.J. and J.M. Robins, *The Identification of Synergism in the Sufficient-Component-Cause Framework*. Epidemiology, 2007. **18**(3): p. 329-339.
4. Lee, W.-C., *Assessing Causal Mechanistic Interactions: A Peril Ratio Index of Synergy Based on Multiplicativity*. PLOS ONE, 2013. **8**(6): p. e67424.
5. Hoffmann, K., C. Heidemann, C. Weikert, M.B. Schulze and H. Boeing, *Estimating the Proportion of Disease due to Classes of Sufficient Causes*. American Journal of Epidemiology, 2005. **163**(1): p. 76-83.
6. VanderWeele Tyler, J., *Attributable Fractions for Sufficient Cause Interactions*, in *The International Journal of Biostatistics*. 2010.
7. Liao, S.-F. and W.-C. Lee, *Weighing the Causal Pies in Case-Control Studies*. Annals of Epidemiology, 2010. **20**(7): p. 568-573.
8. Lee, W.-C., *Completion Potentials of Sufficient Component Causes*. Epidemiology, 2012. **23**(3): p. 446-453.
9. Lin, S.-H., Y.T. Huang and H.-I. Yang, *On identification of agonistic interaction: Hepatitis B and C interaction on hepatocellular carcinoma*. Statistics in Medicine, 2019. **38**.
10. Lin, J.-H., K.-I. Lin, W.-C. Lee and S.-H. Lin, *Stochastic approach for mechanistic interaction under longitudinal studies with non-informative right censoring*. Statistics in Medicine, 2019. **39**.
11. Chen, C.L., H.I. Yang, W.S. Yang, C.J. Liu, P.J. Chen, S.L. You, . . . C.J. Chen, *Metabolic Factors and Risk of Hepatocellular Carcinoma by Chronic Hepatitis B/C Infection: A Follow-up Study in Taiwan*. Gastroenterology, 2008. **135**(1): p. 111-121.
12. Lauer, G.M. and B.D. Walker, *Hepatitis C Virus Infection*. New England Journal of Medicine, 2001. **345**(1): p. 41-52.
13. Lee, M.-H., H.-I. Yang, S.-N. Lu, C.-L. Jen, S.-H. Yeh, C.-J. Liu, . . . C.-J. Chen, *Hepatitis C Virus Seromarkers and Subsequent Risk of Hepatocellular Carcinoma: Long-Term Predictors From a Community-Based Cohort Study*. Journal of Clinical Oncology, 2010. **28**(30): p. 4587-4593.

14. Iloeje, U.H., H.I. Yang, C.L. Jen, J. Su, L.Y. Wang, S.L. You and C.J. Chen, *Risk and Predictors of Mortality Associated With Chronic Hepatitis B Infection*. *Clinical Gastroenterology and Hepatology*, 2007. **5**(8): p. 921-931.
15. Chen, C.-J., H.-I. Yang, J. Su, C.-L. Jen, S.-L. You, S.-N. Lu, . . . f.t. Reveal-Hbv Study Group, *Risk of Hepatocellular Carcinoma Across a Biological Gradient of Serum Hepatitis B Virus DNA Level*. *JAMA*, 2006. **295**(1): p. 65-73.
16. Huang, Y.-T., C.-L. Jen, H.-I. Yang, M.-H. Lee, J. Su, S.-N. Lu, . . . C.-J. Chen, *Lifetime risk and sex difference of hepatocellular carcinoma among patients with chronic hepatitis B and C*. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 2011. **29**(27): p. 3643-3650.
17. Splawa-Neyman, J., D.M. Dabrowska and T.P. Speed, *On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9*. *Statistical Science*, 1990. **5**(4): p. 465-472.
18. Pearl, J. and D. Mackenzie, *The Book of Why: The New Science of Cause and Effect*. 2018: Basic Books, Inc.
19. Brady, H., D. Collier and J. Sekhon, *The Neyman-Rubin Model of Causal Inference and Estimation Via Matching Methods*. *The Oxford Handbook of Political Methodology*, 2008.
20. Rubin, D.B., *Estimating causal effects of treatments in randomized and nonrandomized studies*. *Journal of Educational Psychology*, 1974. **66**(5): p. 688-701.
21. Mutlu, U., M.A. Ikram and M.K. Ikram, *Clinical interpretation of negative mediated interaction*. *International Journal of Epidemiology*, 2018. **48**(4): p. 1286-1293.
22. Flanders, W.D., *On the relationship of sufficient component cause models with potential outcome (counterfactual) models*. *European Journal of Epidemiology*, 2006. **21**(12): p. 847-853.
23. Greenland, S. and C. Poole, *Invariants and noninvariants in the concept of interdependent effects*. *Scandinavian Journal of Work, Environment & Health*, 1988. **14**(2): p. 125-129.
24. VanderWeele, T.J. and M.A. Hernán, *From counterfactuals to sufficient component causes and vice versa*. *European Journal of Epidemiology*, 2006. **21**(12): p. 855-858.
25. Vanderweele, T.J. and J.M. Robins, *Empirical and counterfactual conditions for sufficient cause interactions*. *Biometrika*, 2008. **95**(1): p. 49-61.
26. Suzuki, E., E. Yamamoto and T. Tsuda, *On the Link Between Sufficient-cause Model and Potential-outcome Model*. *Epidemiology*, 2011. **22**(1): p. 131-132.
27. Lee, W.-C., *Testing Synergisms in a No-Redundancy Sufficient-Cause Rate*

- Model*. *Epidemiology*, 2013. **24**(1): p. 174-175.
28. VanderWeele Tyler, J. and J. Knol Mirjam, *A Tutorial on Interaction*, in *Epidemiologic Methods*. 2014. p. 33.
 29. VanderWeele, T.J., J.P. Vandenbroucke, E.J.T. Tchetgen and J.M. Robins, *A mapping between interactions and interference: implications for vaccine trials*. *Epidemiology* (Cambridge, Mass.), 2012. **23**(2): p. 285-292.
 30. Suzuki, E., E. Yamamoto and T. Tsuda, *On the Relations Between Excess Fraction, Attributable Fraction, and Etiologic Fraction*. *American Journal of Epidemiology*, 2012. **175**(6): p. 567-575.
 31. VanderWeele, T.J. and J.M. Robins, *Directed Acyclic Graphs, Sufficient Causes, and the Properties of Conditioning on a Common Effect*. *American Journal of Epidemiology*, 2007. **166**(9): p. 1096-1104.
 32. Gomaa, A.-I., S.-A. Khan, M.-B. Toledano, I. Waked and S.-D. Taylor-Robinson, *Hepatocellular carcinoma: epidemiology, risk factors and pathogenesis*. *World journal of gastroenterology*, 2008. **14**(27): p. 4300-4308.
 33. Efron, B. and C. Stein, *The Jackknife Estimate of Variance*. *The Annals of Statistics*, 1981. **9**(3): p. 586-596.
 34. McIntosh, A.I. *The Jackknife Estimation Method*. 2016.

