

# 生物統計研究專題

李揚<sup>1,2,3</sup> 趙青<sup>4</sup> 馬雙鵠<sup>5,6</sup>

1. 中國人民大學應用統計科學研究中心
2. 中國人民大學統計學院
3. 中國人民大學統計諮詢研究中心
4. Merck Research Laboratories
5. 美國耶魯大學生物統計系
6. 太原理工大學數學學院統計系

\* 本文（“生物統計的研究進展與挑戰”）原刊載於《統計研究》33 卷 6 期，本專題略作修改，已經作者同意刊登於生統 eNews

- 李揚，男，33 歲，2010 年畢業於中國人民大學統計學院，獲經濟學博士學位，現為中國人民大學統計學院副教授，國際統計學會推選會員，中國人民大學統計諮詢研究中心主任，北京生物醫學統計與資料管理研究會副秘書長。研究方向為生物統計、決策與預測。
- 趙青，女，26 歲，2015 年畢業於美國耶魯大學生物統計系，獲生物統計系博士學位，現擔任 Merck Research Laboratories (MRL) 生物統計師。研究方向為藥物研發試驗設計、統計分析和決策分析，以及癌症生物標記研究和疾病預測。
- 馬雙鵠，男，38 歲，2004 年畢業於美國威斯康辛大學統計學系，獲統計學博士學位，現為美國耶魯大學生物統計系終身教授。研究方向為生物統計、衛生經濟學與癌症研究等。

## 【第一部分：背景】

生物統計 (Biostatistics) 研究起源於二十世紀研究者對農業 (漁業) 生產中影響因素效用的討論，並逐漸延伸到以人為核心的生物學、(中西) 醫學、公共衛生學等領域。近年來，隨著人們在醫學、製藥學等領域研究的深入，越來越多的科學問題需要以量化研究依據作為支撐，生物統計學研究得以迅速發展。

近幾十年來，生物統計與其他學科協同促進現代醫療發展的例子層出不窮。根據最新資料統計，西方國家男性患者的肺癌發病率較 1991 年以每年 1.8% 的速度持續下降 (Siegel et al., 2016)。這一方面得益於肺癌預防與早期治療技術的發展。另一方面，大量流行病研究讓公眾意識到吸菸和肺癌之間的關聯，使這一危險因素被有效控制。隨著基因技術的發展，臨床腫瘤試驗被廣泛運用於癌症標靶治療的研究發展中。特羅凱 (鹽酸厄洛替尼片) 是全球首個治療非小細胞肺癌的標靶治療藥物，以人體表皮生長因子受體為靶點。在一項國際大型的隨機、雙盲、安慰劑對照的 III 期臨床試驗中，特羅凱被證實了其顯著的有效性，於 2004 年 11 月獲得美國 FDA 批准上市 (Cohen et al., 2005)。研究者針對這一臨床試驗的主要結局指標進行了存活模型分析，結果發現使用特羅凱的肺癌患者比安慰劑組患者生存期顯著延長 (總生存期中位數延長達 42.5%)。存活分析大大增進了醫學研究者對癌症治療的預後和治療效果的認識。然而，現有的通過表皮生長因子突變的檢測決定的治療方案只對 12% 的患者有效。為了進一步探索疾病機制讓更多患者受益，德州大學安德森癌症中心近期通過基因資料分析和結合生物標記物資訊的創新性臨床試驗找到了另一個重要的致癌基因 (KRAS 基因)。該基因的三個表現型呈現出截然不同的體內信號，為更精準的醫療計畫和標靶藥物的研究策略提供了方向 (Skoulidis et al., 2015)。

作為應用性交叉學科，生物統計研究的重點是理論研究向應用研究的過渡與轉化。近年來，越來越多的研究者將生物統計分析技術廣泛應用於公共衛生、衛生經濟、社會保障等學科的調查與分析，以客觀的定量分析結果為實際工作者的科學研究提供堅實的資料保障，產生的社會影響與效益日益顯著。但是，隨著醫藥衛生行業的不斷細分和複雜化，生物統計的研究對象資料呈現出個體化、高維化、結構化、空間化的特徵。研究方法逐漸由傳統統計學模型延伸到相關結構資料分析、生物資訊學、統計模擬與資料採礦、空間統計分析等統計學前沿領域。

## 【第二部分：流行病學研究】

流行病學以探索暴露因素與疾病間因果關係為目的，起源於古希臘時期。作為公共衛生與醫學研究的基礎方法，其理論框架、研究設計方法在二十世紀八十年代逐漸完善，並在九十年代與相關學科進一步相互滲透發展 (李立明, 2010)。橫斷面研究 (Cross-sectional Study) 是最常用的流行病學研究方法，目前仍常見於各種疾病的流行病學研究 (Souza et al., 2013; Dutra and Glantz, 2014)。雖然沒有考慮暴露因素與疾病關係隨時間影響的非線性變化，但其成本低、易於實施的特性受到廣大研究者的青睞。相應地，病例對照研究 (Case-control Study) 作為臨床流行病學研究中的重要研究方法 (Kotloff et al., 2013; Singh et al., 2013) 以現有疾病患者為病例，以具有可比性的未患病個體為對照，既可以分析暴露因素對疾病狀態的靜態

影響，也可以針對重複觀測資訊分析將疾病變化因素納入研究。但由於病例對照研究在配對過程中可能存在無法控制的干擾因子或選擇性偏倚，其因果關係結論需要研究者審慎對待。為了得到更強的流行病學證據，越來越多的領域 (Pearce et al., 2012; Titulaer et al., 2013) 開始展開大規模世代研究 (Cohort Study)。世代研究以某一特定人群為研究對象，長期採集其暴露因素與疾病狀態資料，為流行病學研究積累了大量時間、空間和其他方面的重要資料。世代研究雖然具有無回憶性偏倚、因果關係推斷證據強等優點，但由於其研究對象的多樣性和資料獲取的豐富性，需要極大的人力、物力支持消耗，因此不宜大規模廣泛使用。

隨著科學研究的進展，流行病學的研究方法被應用到不同的研究領域，並與多學科交叉發展，形成新的研究方向。譬如，為了進一步分析生物標誌物對疾病狀態的影響，研究者以分子生物學技術為手段探索疾病影響因素的分子流行病學 (Wacholder et al., 2004)；為探討歷史、社會、環境等多因素對健康與疾病影響而產生的生態流行病學 (March and Susser, 2006)；考慮生物遺傳多樣性與病原體進化等因素的進化流行病學 (Restif, 2009)；考慮不同地區人口、環境、行為交互影響關係的空間流行病學 (Ostfeld et al., 2005) 等等。上述研究方向分別從宏觀、微觀不同角度展開研究，利用資料資訊提煉支撐公共衛生與醫學研究結論的重要證據。某種程度上，流行病學與生物統計學的研究發展相輔相成。生物統計方法學研究不斷為流行病學研究提供新的方法學保障，流行病學研究為生物統計方法學研究不斷提出新的挑戰：

(1) 在分子流行病學研究中，研究者通過醫學實驗發現潛在的基因影響因素，這些多維 (或高維) 危險/保護因素的資料結構引起了生物統計學界關於降維與變數選擇的研究熱潮 (詳見本文第五部分，於生統 eNews 第?期刊出)。

(2) 由於流行病學研究中資料獲取研究設計與現場實施的限制，可能存在無法控制或觀測的干擾因子。干擾因子的存在嚴重破壞了研究結果中危險因素與疾病狀態因果關係的推斷。上述問題引起了生物統計研究者關於因果推斷 (Casual Inference) 的熱議，衍生出傾向分數配對 (Propensity Score Matching) 等一系列分析方法 (Austin, 2013; Imai and Ratkovic, 2014; Maeda et al., 2016)。

(3) 很多流行病學研究中存在危險因素 (如污染程度、飲食結構等) 無法被直接測量的困境。雖然實證研究中可以採用間接測量作為替代，但由於該方法易在建模分析中引入測量誤差 (Measurement Error) 和錯誤識別 (Misspecification) 問題，成為近年來生物統計模型研究的熱點 (Sapiro and Paciorek, 2013; Szpiro et al., 2014; Alexeeff et al., 2015)。

上述方法的廣泛探索，在豐富生物統計方法的同時提高了流行病學實證研究中模型估計的精度與解釋的力度。

### 【第三部分：臨床試驗設計研究】

2015 年是傳統醫學界振奮人心的一年，屠呦呦研究員憑藉青蒿素的發現獲得該年度諾貝爾生理醫學獎。實踐證明青蒿素及其衍生物青蒿琥酯、蒿甲醚能迅速消滅人體內瘧原蟲，對腦瘧等惡性瘧疾有很好的治療效果。然而一種藥物 (或元素) 從發現到臨床推廣的漫長過程，需要醫學工作者和生物統計研究者付出大量的努力。臨床試驗是驗證藥物安全性和有效性的必要途徑。根據美國國立衛生研究院 (NIH) 的定義，臨床試驗是以人為研究對象通過

問題或測量來驗證生物或行為干預(如藥物、治療、器械等)結果的過程。Ehrhardt 等人(2015)指出 2014 年在 NIH 資料庫中新註冊的臨床試驗共 18,400 項(其中由製藥企業資助的試驗 6,550 項,由 NIH 資助的試驗 1,048 項)。臨床試驗根據其研究目的可分為檢驗安全性的第 I 期試驗、探索有效性的第 II 期試驗、驗證有效性的第 III 期試驗和用於上市後評價的第 IV 期試驗。其中第 III 期臨床試驗由於樣本多、成本高等特點成為監管部門與製藥公司研究和關注的熱點。臨床試驗設計方法是藥物開發過程中的關鍵技術,其主要作用是在藥物安全性和有效性評估中通過隨機化過程來避免選擇偏倚,控制干擾因子,以保證兩組樣本(試驗組和對照組)結果的可比性(Machin and Fayers, 2010)。目前最常用的設計方法是隨機對照試驗(Random Control Trial)。根據分組形式劃分,隨機對照試驗可分為平行對照試驗(Schulz et al., 2010)、交叉對照試驗(Gottlieb et al., 2009)、集群隨機對照試驗(Puffer et al., 2005)和因子隨機對照試驗(Brunoni et al., 2013)等;根據研究目的不同,隨機對照試驗分為優性隨機對照試驗(Manaseki-Holland et al., 2012)、非劣性效隨機對照試驗(Saver et al., 2012)和等效隨機對照試驗(Piaggio et al., 2012)。

雖然在實證研究中可根據研究目的來調整設計思路,但上述傳統隨機對照試驗設計方法仍存在一些缺點。譬如若受試者在影響因素(譬如病程、併發病等)水準上存在較為嚴重的不平衡情況(譬如絕大多數受試者為長期患者,少數受試者為短期患者;又如絕大多數受試者無併發病,只有少數受試者存在某種併發病),傳統的隨機對照試驗在樣本量有限情況下很難保證試驗組與對照組在所有影響因素水準上都可比。另一方面,傳統的隨機對照試驗可能存在倫理方面的風險:研究者在設計試驗時應已有部分證據顯示試驗組治療的有效性,且這種證據隨著試驗的進行會越來越強。但傳統的隨機對照試驗仍機械地保持兩組各 50% 的概率納入受試者,使得對照組的受試者面臨更大的感染風險。這樣的做法雖然沒有破壞倫理的規則,但仍在一定程度上存在倫理的風險。因此,能否在保證隨機性的基礎上對傳統方法進行改進,使其更好地利用試驗過程中的資訊,更好地保證試驗組與對照組的可比性,是調整(adaptive)隨機對照試驗產生的源動力。調整隨機對照試驗根據上一個受試者的分組或干預結果,按某一特定隨機規則確定下一個受試者進入某一組(試驗組/對照組)的機率(Chow & Chang, 2008),既保證樣本分配的隨機性,又保證統計分析的有效性。隨著調整隨機對照試驗的研究進展,美國食品與藥品管理局(FDA)從 2010 年起開始制定關於調整臨床試驗設計方法的監管條例指南(Rosenberger et al., 2012)。另一方面,由於調整隨機對照試驗可以針對不同表現型患者展開有針對性的隨機化模式,因此可以用於探索同一藥物在不同類型患者間的適用性。因此,在美國廣大研究者將調整設計看做解決歐巴馬政府 2015 年 1 月發起總預算 2015 億美元的精準醫療計畫的一個重要思路。調整臨床試驗設計主要分為限制隨機化方法(Restricted randomization)、反應調整隨機分派(Response-adaptive randomization)、共變項調整隨機分派(Covariate-adaptive randomization)三種主要類型。

限制隨機方法可以看作完全隨機方法(Complete randomization)的直接拓展,通過調整隨機分派過程的分組機率,保證試驗在可比性(Balance)和隨機性(Randomization)二者間達到較好的平衡(Rosenberger and Lachin, 2002)。反應調整隨機分派方法在限制隨機方法的基礎上進一步考慮醫學研究的倫理性,即在不破壞試驗隨機性的前提下儘量將受試者分配

到更有效的治療組中。在 Efron (1971) 的有偏硬幣隨機試驗 (Biased coin designs, BCD) 基礎上, Markaryan 和 Rosenberger (2010) 通過理論分析給出了該方法的精確性質。延續這一思想, Baldi 等 (2004; 2008) 提出了對小樣本情況也具有較好檢驗功效的調整有偏硬幣隨機試驗 (Adjusted BCD), Hu 等 (2004; 2009) 提出了漸進性質最優 (分配比例變異性最小) 的雙重有偏硬幣隨機試驗 (Doubly biased coin designs, DBCD) 並證明了相關理論性質。當受試者在影響因素水準上不平衡且樣本量有限時, 單純的隨機化方法不能保證試驗組和對照組受試者的可比性。共變數調整隨機分派方法針對這一情況採用分層隨機 (Stratified Randomization)、邊際共變數調整隨機分派 (Marginal Covariate-adaptive randomization) 或同質線性模型最優設計過程 (Optimal designs for homoscedastic linear models) 的思路 (Brown et al., 2005; Shao et al., 2010; Baldi et al., 2011; Yuan et al., 2011)。Ma 等人 (2015) 結合上述思路提出一個最優化調整過程並證明其在統計推斷方面的優良性質。需要注意的是, 上述方法僅從一個角度討論調整隨機分派試驗的設計方法。實際問題中往往存在更為複雜的情況, 需要展開不同設計方法的交叉研究。譬如反應變數調整共變數調整隨機分派方法 (Zhu and Hu, 2010; Hu et al., 2015)、縱向反應變數調整共變數調整隨機分派方法 (Huang and Zhu, 2016)、非劣性調整隨機分派試驗 (Kwong et al., 2013; Hartley, 2015) 等等。

臨床試驗設計方法的研究還面臨很多挑戰:

(1) 首先, 雖然諸多生物統計研究者探索了適用於不同研究目的的隨機化方法, 但這些方法由於在隨機過程中做了一定調整與限定, 其資料結果在因果推斷層面勢必受到一定影響, 需要廣大研究者在設計隨機化過程時考慮資料分析層面的不偏性與有效性。

(2) 其次, 隨機化方法的理論假設與臨床操作間可能存在一定差距, 需要研究者進一步討論。譬如反應變數調整隨機分派方法要求試驗結果“立即可觀察”, 很難在實踐中保證。需要研究者結合操作實際進一步方法的適用性及其理論性質。

(3) 第三, 醫學真實世界的複雜性導致人群的高維異質性特點。為保證精準醫療的有效調整設計, 如何在高維度因子數據中選擇有效的共變數進行隨機化程序控制, 成為即待解決的科學問題。

～～ 第四部分：存活分析研究、第五部份：基因資料分析研究 ～～

～～ 下期待續：(2016/12/28)出刊 ～～

## 參考文獻

1. Siegel, R. L., Miller, K. D., & Jemal, A. (2016). *Cancer statistics, 2016*. *CA: a cancer journal for clinicians*, 66(1), 7-30.
2. Cohen, M. H., Johnson, J. R., Chen, Y. F., Sridhara, R., & Pazdur, R. (2005). *FDA drug approval summary: erlotinib (Tarceva®) tablets*. *The oncologist*, 10(7), 461-466.
3. Skoulidis, F., Byers, L. A., Diao, L., Papadimitrakopoulou, V. A., Tong, P., Izzo, J., ... & Zhang, J. (2015). *Co-occurring genomic alterations define major subsets of KRAS-mutant lung adenocarcinoma with distinct biology, immune profiles, and therapeutic vulnerabilities*. *Cancer discovery*, 5(8), 860-877.
4. Machin, D., Fayers, P. (2010). *Randomized clinical trials: design, practice and reporting*. John Wiley & Sons, Oxford.
5. Schulz, K.F., Altman, D.G., & Moher, D. (2010). *CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials*. *Annals of internal medicine*, 152(11), 726-732.
6. Gottlieb, A., Menter, A., Mendelsohn, A., Shen, Y.K., Li, S., Guzzo, C., Kavanaugh, A. (2009). *Ustekinumab, a human interleukin 12/23 monoclonal antibody, for psoriatic arthritis: randomised, double-blind, placebo-controlled, crossover trial*. *The Lancet*, 373(9664), 633-640.
7. Puffer, S., Torgerson, D.J., & Watson, J. (2005). *Cluster randomized controlled trials*. *Journal of Evaluation in Clinical Practice*, 11(5), 479-483.
8. Brunoni, A. R., Valiengo, L., Baccaro, A., Zanao, T. A., de Oliveira, J. F., Goulart, A., Fregni, F. (2013). *The sertraline vs electrical current therapy for treating depression clinical study: results from a factorial, randomized, controlled trial*. *JAMA psychiatry*, 70(4), 383-391.
9. Manaseki-Holland, S., Maroof, Z., Bruce, J., Mughal, M. Z., Masher, M. I., Bhutta, Z. A., Chandramohan, D. (2012). *Effect on the incidence of pneumonia of vitamin D supplementation by quarterly bolus dose to infants in Kabul: a randomised controlled superiority trial*. *The Lancet*, 379(9824), 1419-1427.
10. Saver, J. L., Jahan, R., Levy, E. I., Jovin, T. G., Baxter, B., Nogueira, R. G., Zaidat, O. O. (2012). *Solitaire flow restoration device versus the Merci Retriever in patients with acute ischaemic stroke (SWIFT): a randomised, parallel-group, non-inferiority trial*. *The Lancet*, 380(9849), 1241-1249.
11. Piaggio, G., Elbourne, D. R., Pocock, S. J., Evans, S. J., Altman, D. G., & CONSORT Group. (2012). *Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement*. *The Journal of the American Medical Association*, 308(24), 2594-2604.
12. Chow, S. C., & Chang, M. (2008). *Adaptive design methods in clinical trials-a review*. *Orphanet Journal of Rare Diseases*, 3(11), 169-90.
13. Rosenberger, W. F., Sverdlov, O., & Hu, F. (2012). *Adaptive randomization for clinical trials*.

*Journal of biopharmaceutical Statistics*, 22(4), 719-736.

14. Rosenberger, W. F., Lachin, J. L. (2002). *Randomization in Clinical Trials: Theory and Practice*. New York: Wiley.
15. Ehrhardt, S., Appel, L. J., & Meinert, C. L. (2015). Trends in National Institutes of Health Funding for Clinical Trials Registered in ClinicalTrials.gov. *Journal of the American Medical Association*, 314(23), 2566-2567.
16. Hu, J., Zhu, H., & Hu, F. (2015). A unified family of covariate-adjusted response-adaptive designs based on efficiency and ethics. *Journal of the American Statistical Association*, 110(509), 357-367.
17. Huang, T., & Zhu, H. (2016). Longitudinal Covariate-Adjusted Response-Adaptive Randomization: Impact of Missing Data. *Modern Adaptive Randomized Clinical Trials: Statistical and Practical Aspects*, 81, 327. Boca Raton: CRC Press.
18. Kwong, K. S., Cheung, S. H., & Hayter, A. J. (2013). Step-up procedures for non-inferiority tests with multiple experimental treatments. *Statistical methods in medical research*, 0962280213477767.
19. Hartley, A. M. (2015). A Bayesian adaptive blinded sample size adjustment method for risk differences. *Pharmaceutical Statistics*, 14(6), 488-514.
20. Souza, J. P., Gülmezoglu, A. M., Vogel, J., Carroli, G., Lumbiganon, P., Qureshi, Z., ...& Neves, I. (2013). Moving beyond essential interventions for reduction of maternal mortality (the WHO Multicountry Survey on Maternal and Newborn Health): a cross-sectional study. *The Lancet*, 381(9879), 1747-1755.
21. Dutra, L. M., & Glantz, S. A. (2014). Electronic cigarettes and conventional cigarette use among US adolescents: a cross-sectional study. *JAMA pediatrics*, 168(7), 610-617.
22. Kotloff, K. L., Nataro, J. P., Blackwelder, W. C., Nasrin, D., Farag, T. H., Panchalingam, S., ...& Faruque, A. S. (2013). Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *The Lancet*, 382(9888), 209-222.
23. Singh, S., Chang, H. Y., Richards, T. M., Weiner, J. P., Clark, J. M., & Segal, J. B. (2013). Glucagonlike peptide 1-based therapies and risk of hospitalization for acute pancreatitis in type 2 diabetes mellitus: a population-based matched case-control study. *JAMA internal medicine*, 173(7), 534-539.
24. Pearce, M. S., Salotti, J. A., Little, M. P., McHugh, K., Lee, C., Kim, K. P., ...& Parker, L. (2012). Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study. *The Lancet*, 380(9840), 499-505.
25. Titulaer, M. J., McCracken, L., Gabilondo, I., Armangué, T., Glaser, C., Iizuka, T., ...& Aguilar, E. (2013). Treatment and prognostic factors for long-term outcome in patients with anti-NMDA receptor encephalitis: an observational cohort study. *The Lancet Neurology*, 12(2), 157-165.
26. March, D., & Susser, E. (2006). The eco-in eco-epidemiology. *International Journal of*

*Epidemiology*, 35(6), 1379-1383.

27. Wacholder, S., Chanock, S., Garcia-Closas, M., & Rothman, N. (2004). Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *Journal of the National Cancer Institute*, 96(6), 434-442.
28. Restif, O. (2009). Evolutionary epidemiology 20 years on: challenges and prospects. *Infection, Genetics and Evolution*, 9(1), 108-123.
29. Ostfeld, R. S., Glass, G. E., & Keesing, F. (2005). Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends in ecology & evolution*, 20(6), 328-336.
30. Austin, P. C. (2013). The performance of different propensity score methods for estimating marginal hazard ratios. *Statistics in medicine*, 32(16), 2837-2849.
31. Imai, K., & Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1), 243-263.
32. Maeda, I., Morita, T., Yamaguchi, T., Inoue, S., Ikenaga, M., Matsumoto, Y., ...& Tatara, R. (2016). Effect of continuous deep sedation on survival in patients with advanced cancer (J-Proval): a propensity score-weighted analysis of a prospective cohort study. *The Lancet Oncology*, 17(1), 115-122.
33. Szpiro, A. A., & Paciorek, C. J. (2013). Measurement error in two-stage analyses, with application to air pollution epidemiology. *Environmetrics*, 24(8), 501-517.
34. Alexeeff, S. E., Carroll, R. J., & Coull, B. (2015). Spatial measurement error and correction by spatial SIMEX in linear regression models when using predicted air pollution exposures. *Biostatistics*, kxv048.
35. Szpiro, A. A., Sheppard, L., Adar, S. D., & Kaufman, J. D. (2014). Estimating acute air pollution health effects from cohort study data. *Biometrics*, 70(1), 164-174.