



# History of the modern epidemiological concept of confounding

Alfredo Morabia

## Correspondence to

Alfredo Morabia, Center for the Biology of Natural Systems, Queens College - CUNY, 163-03 Horace Harding Expressway, Flushing, NY 11365, USA; [alfredo.morabia@qc.cuny.edu](mailto:alfredo.morabia@qc.cuny.edu)

Accepted 19 March 2010  
Published Online First  
9 August 2010

## ABSTRACT

The epidemiological concept of confounding has had a convoluted history. It was first expressed as an issue of group non-comparability, later as an uncontrolled fallacy, then as a controllable fallacy named confounding, and, more recently, as an issue of group non-comparability in the distribution of potential outcome types. This latest development synthesised the apparent disconnect between phases of the history of confounding. Group non-comparability is the essence of confounding, and the statistical fallacy its consequence. This essay discusses how confounding was perceived in the 18th and 19th centuries, reviews how the concept evolved across the 20th century and finally describes the modern definition of confounding.

## INTRODUCTION

To an unprepared mind, the terms 'confounding' or 'confounder' do not immediately evoke the consequences of comparing groups of people who differ on determinants of the studied outcome. Expressions suggesting the imbalanced distribution of multiple independent causes across groups would have conveyed the meaning more directly, but epidemiology retained the verb confound. The reason is that the theoretical work on the concept of confounding started with a description of a statistical fallacy: under some conditions, the effect of an exposure could be similar in each stratum of a third variable, but when these strata were pooled, it was as if the effect of the exposure of interest got 'mixed' with that of the third variable.<sup>1</sup> The fallacy was thus aptly named confounding, from an old usage of the Medieval Latin verb 'confundere', which meant mixing.<sup>2</sup> Earlier attempts to trace the history of confounding essentially focused on this conceptualisation of confounding as a fallacy.<sup>2-3</sup>

Compared to earlier reports, the present essay expands the history of the concept known today as epidemiologic confounding in the phases preceding and following the time when it was primarily viewed as a fallacy. After discussing how confounding was perceived in the 18th and 19th centuries, the essay reviews how the concept evolved across the 20th century and finally describes the modern definition of confounding.

## METHOD

The methodological approach driving this history of confounding is inspired by Piaget's genetic epistemology.<sup>3-4</sup> The leading idea is that scientific disciplines are in continual construction, formalisation and organisation. Their methods and concepts are commonsensical when the discipline

first appears, but become increasingly theoretical and abstract as the discipline acquires experience and addresses questions of increasingly complex nature.<sup>4</sup>

The genetic epistemology approach assumes that the concept that was eventually named confounding: A. had a history, which started with commonsensical observations, B. evolved into an increasingly abstract, formal and overarching concept, and C. is still evolving today.

To trace the history of confounding, this essay uses the four phases (preformal, early, classic and modern) previously identified in the history of epidemiological methods and concepts.<sup>3</sup> It also focuses on the theory of confounding, and does not cover the statistical approaches to confounding-related issues, (eg, collapsibility<sup>5</sup>) methods of adjusting for confounding, or the history of causal inference, which, even though closely related and overlapping at times with that of confounding, has a broader scope.<sup>5-6</sup>

## PREFORMAL CONFOUNDING

Non-comparability of groups is the most primitive epidemiological concern to which the modern concept of confounding can be traced. When, in 1747, Lind<sup>7</sup> compared the efficacy of candidate treatments of scurvy, he made sure his six experimental pairs of seamen were comparable, a priori, in terms of determinants of scurvy lethality such as disease stage, food and air quality.<sup>8</sup>

In the 19th century, group non-comparability was a formidable criticism to epidemiological studies.<sup>9</sup> Hence the emphasis put by John Snow<sup>10</sup> on the comparability of the 1854 London clients of the Southwark and Vauxhall water company, who drank polluted Thames water and experienced high mortality from cholera, with those of the Lambeth Company, who received relatively sewage-free water and experienced low mortality from cholera. Both groups, Snow insisted, were similar in social standing, housing space and occupations. He specifically investigated neighbourhoods with mixed water supply, in which adjacent houses could be supplied by different water companies.<sup>11</sup> But for Snow's contemporaries, like Farr<sup>12</sup>, who believed that cholera was due to air pollution, that is, miasma,<sup>13</sup> the two companies served clients who differed substantially in ways thought to be relevant to the occurrence of cholera, such as elevation above sea level, family income and quality of housing. How could Snow confound his critics? He could only speculate that the clients of the two companies must have been comparable as a large number of people ('no fewer than 300 000'), 'were divided into two groups without their choice, and,

in most cases, without their knowledge'.<sup>11</sup> Retrospectively, we understand that the contagionist Snow was arguing against the miasmatic idea that the two client populations were comparable on some miasma-related confounding characteristics. Snow claimed that comparability was plausible, but he lacked the techniques developed subsequently to achieve comparability analytically or by design.

**EARLY THEORY OF CONFOUNDING**

Indeed, epidemiologists of the first half of the 20th century began to formally address the criticism of non-comparability.<sup>9</sup> They implemented new techniques such as random allocation of treatment<sup>14</sup>, restriction of the study sample<sup>15</sup>, standardisation of risks and rates<sup>15 16</sup> and exposure propensity scores.<sup>17 18</sup> They improved the epidemiological study designs, such as retrospective cohort studies<sup>16 19</sup> and case—control studies.<sup>17 20</sup> All of these efforts aimed to design studies and/or analyse the data in ways that purposefully optimised comparability on alternative causes of the studied outcome. Surprisingly, the first definition of the concept we refer to today as 'confounding,' did not follow from this line of efforts to achieve balanced comparisons.

As shown in table 1, Yule,<sup>25</sup> in 1903,<sup>1</sup> Greenwood,<sup>26</sup> in 1935,<sup>21</sup> and Hill,<sup>27</sup> in 1939,<sup>22</sup> apparently independently described a fallacy resulting from pooling data when a third variable was not equally distributed in the compared groups. Yule used the imaginary example of an attribute, not transmitted by fathers to sons or by mothers to daughters, but that showed 'considerable apparent inheritance' when the data of fathers, sons, mothers and daughters were analysed together. Greenwood,<sup>21</sup> imagined an immunisation experiment, in which risk of death was similar among the inoculated and the non-inoculated in first and second groups of patients, but pooling the two groups together resulted in a spurious protective effect of the inoculation. In Hill's example, a treatment did not work for men or women, but reduced mortality when the male and female data were combined.<sup>22</sup>

Yet, Yule, Greenwood and Hill do not seem to have viewed the fallacy as a common issue in population studies, and did not suggest computing a weighted average of the stratum-specific effects to bypass it. Apparently, their examples went into oblivion. The subsequent phase of the history of the epidemiological concept of 'confounding' appears to be an offshoot of discussions related to the modelling of interactions.

**CLASSIC THEORY OF CONFOUNDING**

Fisher<sup>28</sup> used the verb 'confound' in 1926, to describe the implication of discarding some high-order interactions in the analysis of data from studies with factorial designs.<sup>2</sup> Precision could be improved, but the sacrifice of interactions would

amalgamate strata, eliminating and therefore 'confounding' the manifestation of some of the underlying heterogeneity of effects.<sup>29</sup>

In my view, Fisher was using the term 'confounding' in the same way it had been used earlier by the English philosopher Mill, that is, as the consequence of ignoring causal interactions. For Mill, confounding meant 'intermixture of causes,' which he defined as two or more causes, 'modifying the effects of one another'.<sup>30</sup> Mill was referring to a mixing of effects that were heterogeneous across strata of one of the causes. This was different from Yule's fallacy, in which the exposure had a single effect, which was similar in all strata of the extraneous factor, except for being confounded in the pooled effect.

It is at that point that Simpson,<sup>31</sup> building on Fisher's work, made the contribution now known as Simpson's paradox. Simpson showed that discarding the interaction terms could impact the estimation of the pooled effect even when the stratum-specific effects were homogeneous. This could actually leave 'considerable scope for paradox and error'. He gave the example of an imaginary trial, (see table 1) in which the treatment homogeneously increased the survival odds both for males and females as separate groups, but had no effect when genders were pooled.<sup>23</sup>

Simpson posited that, for second-order interactions to be ignored, the third variable had to be independent of the treatment among the non-outcome and independent of the outcome variable among the unexposed. Otherwise, stratification had to be preserved. This became the core of the classic epidemiological definition of confounding.

From 1959 on, expressions appear in the epidemiology literature, which evoke Yule's fallacy or Simpson's paradox without explicitly referring to them. Papers and textbooks mention 'indirect associations',<sup>32</sup> and 'misleading associations,' produced by 'extraneous factors',<sup>33</sup> or, 'indirect associations generated by factors related to both outcome and exposure'.<sup>34</sup> The term 'confounding' itself began to appear in epidemiological articles and textbooks in the 1970s.<sup>35–37</sup> Its usage may have reflected the influence of the sociologist Kish who had defined the term in 1959.<sup>2 38</sup>

Around 1980, it was specified, in addition to the two conditions formulated by Simpson, that the third variable should not mediate the relation of exposure to outcome.<sup>39 40</sup> This third condition highlighted the need for a priori, non-statistical knowledge about the relationship of the potential confounder with the other studied variables.<sup>41</sup>

Overall, table 1 shows the similarity of the quantitative examples used to illustrate confounding as a mixing of effects, from Yule<sup>1</sup> to Rothman,<sup>24</sup> that is, across most of the 20th century. Yule's expression of confounding as, 'a fallacy caused by the mixing of records (ie, strata)',<sup>1</sup> is analogous to Rothman's,

**Table 1** Five historical representations of the concept of confounding as a fallacy resulting from mixing strata of exposure to a third factor

Author	Contrast	Stratum 1			Stratum 2			ALL		
		Exposure 1	Exposure 2	OR*	Exposure 1	Exposure 2	OR*	Exposure 1	Exposure 2	OR*
Yule <sup>1,†</sup>	Attribute present/absent	0.25/0.25	0.25/0.25	1.0	0.1/0.9	0.1/0.9	1.0	0.13/0.17	0.17/0.53	2.4
Greenwood <sup>21,‡</sup>	Dead/alive	50/50	500/500	1.0	50/950	5/95	1.0	100/1000	505/595	1.0
Hill <sup>22,§</sup>	Dead/alive	16/64	6/24	1.0	16/24	24/36	1.0	32/88	30/60	0.7
Simpson <sup>23,§</sup>	Dead/alive	5/8	3/4	0.8	15/12	3/2	0.8	20/20	6/6	1.0
Rothman <sup>24,¶</sup>	Fit/not fit	9/1	17/3	1.6	3/17	1/9	1.6	12/18	18/12	0.4

\*OR, computed by AM, not in the original publications.

†Stratum 1=fathers; stratum 2=mothers; exposure 1=attribute present; exposure 2=attribute absent.

‡Stratum 1=group 1; stratum 2=group 2; exposure 1=inoculated; exposure 2=not inoculated.

§Stratum 1=males; stratum 2=females; exposure 1=treatment; exposure 2=no treatment.

¶Stratum 1=table 1; stratum 2=table 2; exposure 1=black hats; exposure 2=grey hats.

'on the simplest level, confounding may be considered as a mixing of effects'.<sup>24</sup>

**MODERN THEORY OF CONFOUNDING**

The classic definition of confounding had weaknesses. It was derived from the relation of additional variables to exposure and outcome, and not from the characteristics of the studied association, such as non-comparability. A variable could meet the classic definition and not be a confounder.<sup>42</sup> Matching for a confounder had different implications in cohort and case-control studies.<sup>39 43</sup> Screening for confounding by comparing the stratum-specific and the pooled effects could lead to different conclusions based on whether one used risk ratios, risk differences or ORs.<sup>39</sup>

The modern definition of confounding was inspired by work in the analysis of randomised controlled trials. In 1923, Neyman<sup>44</sup> defined a causal effect as the impossible contrast between the outcome of a single unit, say an individual, if assigned the experimental treatment, and the outcome of that same individual if concurrently assigned the reference treatment.<sup>45</sup> In 1974, Rubin stated the fundamental problem of effect identification in terms similar to those of Neyman.<sup>46</sup> If 'y(E)-y(C)' is the effect of treatment E versus control C on outcome Y, and assuming y(E) and y(C) need to be measured at time 2 on the same person: 'The problem in measuring y(E)-y(C) is that we can never observe both y(E) and y(C) since we cannot return to time t<sub>1</sub> to give the other treatment'.<sup>46</sup>

Each individual can be observed in only one treatment state at any point in time. Of the two potential outcomes (ie, under the experimental or under the reference treatments), one is observed, and the other needs to remain hypothetical. Thus, as described by Copas in 1973,<sup>47</sup> there could be four individual types of potential outcome pairs for a dichotomous treatment (A and B) followed by a dichotomous outcome (success or failure) according to whether a subject would respond to A and B, A but not B, B but not A, or neither A nor B.

There is literary evidence of the ongoing epidemiological reflection about potential outcomes in the 1980s,<sup>48 49</sup> but it wasn't until a 1986 paper by Greenland and Robins that the potential outcome approach to confounding was made widely accessible to epidemiologists.<sup>42</sup> In Greenland and Robins' paper, the potential outcome model was confined to deterministic risks (ie, risks that can equal either 0 or 1) but it differed from previous discussions<sup>46 50</sup> because—as shown in table 2, which imitates a table in Greenland and Robins' 1986 paper—it used the four 'causal' types<sup>47</sup> dubbed 'doomed', 'exposure causative', 'exposure preventive' and 'immune'.

**Table 2** Definition and notation of potential outcome types and their outcomes according to two potential outcomes

Types	n	Description	Outcome	
			Vaccinated	Not vaccinated
Doomed	d	Does not respond to vaccine	Flu	Flu
Causative	c	Vaccine induces flu	Flu	No flu
Preventive	p	Responds to vaccine	No flu	Flu
Immune	i	Insensitive to flu	No flu	No flu

The example is from a centenarian lady who gets an anti-flu vaccine. When later exposed to the flu, she may develop flu symptoms anyway, as if she had not been vaccinated, because her immune system is too weak to respond to the vaccine. In this case she would be 'doomed' to get the flu. Or she may develop flu symptoms she would otherwise not have had because the vaccine contained a live virus that made her sick. In this case, it is because she is susceptible to a 'causative' vaccine. On the other hand, she may not develop flu because the vaccine protected her (susceptible to the 'preventive' vaccine) or because she has a genetic trait making her immune to influenza viruses ('immune' to flu).

The example in table 2 shows that if a centenarian lady has been vaccinated and does not get the flu, she has no way of knowing whether she was susceptible and the vaccine was 'preventive', or whether she is naturally 'immune'. Similarly, if a non-vaccinated person does not get the flu, she cannot know whether she would have avoided the flu had she been vaccinated. She could be 'doomed' or she could lack the protection of the 'preventive' vaccine. The effect of the vaccine cannot be identified, or its parameter estimated, without knowing both potential outcomes, under vaccination, as well as under no vaccination. This is the logical impasse mentioned by Rubin<sup>46</sup>: both potential outcomes cannot be observed simultaneously in the same person.

Consider now, two large randomised groups of N subjects each, and that in each group, the N subjects are d doomed, c causative, p preventive and i immune to flu, where d+c+p+i=N. One group gets the vaccine and the other does not. The risk difference of getting the flu is identifiable as groups are large and comparable with respect to their potential outcome types, assuming there were no gross violations of the assignment protocol,<sup>51</sup> misclassification, or losses in the follow-up. They are, in Greenland and Robins' terminology, 'exchangeable'. The risk of flu is R<sub>V</sub>=(d+c)/N in those vaccinated, and R<sub>NV</sub>=(d+p)/N in those not vaccinated. The risk difference between the non-vaccinated and the vaccinated is RD=R<sub>V</sub>-R<sub>NV</sub>=(p-c)/N. The risk difference only 'partially identifies' the vaccine effect, because a zero effect could be due to the vaccine causing as many flu cases (c) as it prevents (p). 'Full identification' is possible if, for example, the vaccine does not contain killed or weakened influenza virus, but only split particles of the flu virus, which cannot cause flu. Under this scenario, there are no 'c' subjects and the risk difference is simply (p/N), that is, if c=0.

However, if the groups were not at least 'partially' exchangeable, as if, for example, there were more 'doomed' (eg, centenarians with lethargic immune response) in the vaccinated group than in the non-vaccinated group, the ds would not cancel out, and the risk difference would be confounded.

**What is already known on this subject**

- ▶ Earlier attempts to trace the history of confounding focused on the period when confounding was conceptualised as a fallacy resulting from mixing the effect of the studied variables with that of a third variable.

**What this study adds**

- ▶ The present essay expands the history of the concept known today as epidemiologic confounding to the 18th and 19th century when it began to be viewed as an issue of non-comparability between groups.
- ▶ It also explains how the modern definition of confounding based on potential outcome contrasts has reinstated group non-comparability as the essence of confounding and established that the statistical fallacy, from which confounding draws its name, is a consequence of group non-comparability.

This theory of confounding derived from potential outcome contrasts has been generalised from randomised to observational studies,<sup>52</sup> it has helped to formally distinguish confounding from selection bias<sup>53</sup> and has recently been revisited by its authors.<sup>54</sup>

## CONCLUSION

From a broad historical perspective, the modern definition of confounding based on potential outcome contrasts has reinstated group non-comparability as the essence of confounding, establishing the statistical fallacy as one of its consequences.

**Acknowledgements** This essay was presented as an invited lecture at XVIII IEA World Congress of Epidemiology and the VII Brazilian Congress of Epidemiology, in Porto Alegre, 21–24 September, 2008. I am indebted to Sander Greenland, Sharon Schwartz, Olli Miettinen, Jan Vandenbroucke, Jamey Robins, Timothy Lash, Paolo Vineis and Raj Bhopal for discussions and comments on the many earlier versions of this manuscript.

**Competing interests** None.

**Patient consent** Obtained.

**Provenance and peer review** Not commissioned; externally peer reviewed.

## REFERENCES

1. Yule GU. Notes on the theory of association of attributes in statistics. *Biometrika* 1903;**2**:121–34.
2. Vandenbroucke JP. The history of confounding. In: Morabia A, ed. *History of epidemiological methods and concepts*. Basel: Birkhäuser, 2004:313–26.
3. Morabia A. Epidemiology: an epistemological perspective. In: Morabia A, ed. *History of epidemiological methods and concepts*. Basel: Birkhäuser, 2004:1–126.
4. Piaget J. *Genetic epistemology*. New York: Columbia University Press, 1970.
5. Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. *Stat Sci* 1999;**14**:29–46.
6. Vineis P. Causality in epidemiology. In: Morabia A, ed. *History of epidemiological methods and concepts*. Basel: Birkhäuser, 2004:337–50.
7. Tröhler U. James Lind and scurvy: 1747 to 1795. The James Lind Library, 2003. <http://www.jameslindlibrary.org> (accessed 11 Aug 2008).
8. Lind J. *A treatise of scurvy, 1753*. Edinburgh: University Press, 1953.
9. Morabia A. *Epidemiological methods and concepts in the nineteenth century and their influences on the twentieth century*. In: Holland WW, Olsen J, Florey C, eds. *The development of modern epidemiology. Personal reports from those who were there*. New York: Oxford University Press, 2007:17–29.
10. Frerichs RR. John Snow (1813–1858), 2009. <http://www.ph.ucla.edu/epi/snow/encyclopediasummaryfrerichs.html> (accessed 2 Mar 2010).
11. Snow J. *On the mode of communication of cholera*. 2nd edn. London: Churchill, 1855:74–5.
12. Halliday S. William Farr: Campaigning statistician. *J Med Biogr* 2000;**8**:220–7. [http://www.ph.ucla.edu/epi/snow/farr/farr\\_compiler\\_a.html](http://www.ph.ucla.edu/epi/snow/farr/farr_compiler_a.html).
13. Vinten-Johansen P, Brody H, Paneth N, et al. *Cholera, chloroform and the science of medicine: a life of John Snow*. Oxford: Oxford University Press, 2003:260.
14. Therapeutic Trial Committee of the Medical Research Council. The serum treatment of lobar pneumonia. *Lancet* 1934;**1**:290–5.
15. Goldberger J, Wheeler GA, Sydenstricker E. A study of the relation of family income and other economic factors to pellagra incidence in seven cotton-mill villages of South Carolina in 1916. *Public Health Rep* 1920;**35**:2673–714.
16. Weinberg W. *Die Kinder der Tuberkulösen*. Leipzig: S Hirzel, 1913.
17. Lane-Clayton J. *A further report on cancer of the breast: reports on public health and medical subjects*. London, UK: His Majesty's Stationary Office, 1926. Report No.: 32.
18. Morabia A. Janet Lane-Clayton—interphase epitome. *Epidemiology* 2010;**21**:573–6.
19. Winkelstein W Jr. Vignettes of the history of epidemiology: three firsts by Janet Elizabeth Lane-Clayton. *Am J Epidemiol* 2004;**160**:97–101.

20. Stocks P, Karn M. A co-operative study of the habits, home life, dietary and family histories of 450 cancer patients and of an equal number of control patients. *Ann Eugen* 1933;**5**:237–79.
21. Greenwood M. *Epidemics & crowd diseases: Introduction to the study of epidemiology*. North Stratford, NH: Ayer Company Publishers, Incorporated, 1935: 84–5.
22. Hill AB. *Principles of medical statistics*. London: The Lancet Ltd, 1939:125–7.
23. Simpson EH. The interpretation of interaction in contingency tables. *J Roy Stat Soc Ser B* 1951;**13**:238–41.
24. Rothman KJ. *Modern epidemiology*. 1st edn. Boston: Little Brown and Co, 1986:89.
25. Kendall M. George Udny Yule 1871–1951. *J R Stat Soc* 1952;**115**:156–61.
26. Wilkinson L. 'Greenwood, Major (1880–1949)', Oxford Dictionary of National Biography, 2004. <http://www.oxforddnb.com/view/article/51797> (accessed 26 Feb 2010).
27. Doll R. Austin Bradford Hill. 8 July 1897–18 April 1991. *Biogr Mem Fellows R Soc* 1994;**40**:129–40.
28. Yates F, Mather K. Ronald Aylmer Fisher 1890–1962. *Biogr Mem Fellows R Soc* 1963;**9**:91–120.
29. Street DJ. Fisher's Contributions to Agricultural Statistics. *Biometrics* 1990;**46**:937–45.
30. Mill JS. *A system of logic (Eight Edition 1881)*. In: Nagel E, ed. *J.S. Mill's Philosophy of Scientific Method*. New York: Hafner Publishing Co, 1881:3–356.
31. Edward H Simpson (born 1922). [http://en.wikipedia.org/wiki/Edward\\_H.\\_Simpson](http://en.wikipedia.org/wiki/Edward_H._Simpson), 2010 (accessed 2 Mar 2010).
32. Cornfield J, Haenszel W, Hammond EC, et al. Smoking and lung cancer: recent evidence and a discussion of some questions. *J Natl Cancer Inst* 1959;**22**:173–203.
33. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959;**22**:719–48.
34. MacMahon B, Pugh TF, Ipsen J. *Epidemiologic methods*. Boston: Little, Brown and Co, 1960:14–16.
35. MacMahon B, Pugh TF. *Epidemiology—principles and methods*. Boston: Little, Brown and Co, 1970.
36. Miettinen OS. Matching and design efficiency in retrospective studies. *Am J Epidemiol* 1970;**91**:111–18.
37. Susser M. *Causal thinking in the health sciences*. New York: Oxford, 1973.
38. Kish L. Some statistical problems in research design. *Am Sociol Rev* 1959;**24**:328–38.
39. Miettinen OS, Cook EF. Confounding: essence and detection. *Am J Epidemiol* 1981;**114**:593–603.
40. Greenland S, Neutra R. Control of confounding in the assessment of medical technology. *Int J Epidemiol* 1980;**9**:361–7.
41. Rothman KJ. Epidemiologic methods in clinical trials. *Cancer* 1977;**39**:1771–5.
42. Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. *Int J Epidemiol* 1986;**15**:413–19.
43. Kupper LL, Karon JM, Kleinbaum DG, et al. Matching in epidemiologic studies: validity and efficiency considerations. *Biometrics* 1981;**37**:271–91.
44. O'Connor J, Robertson E. Jerzy Neyman (1894–1981), 2003. <http://www-history.mcs.st-andrews.ac.uk/Biographies/Neyman.html> (accessed 2 Mar 2010).
45. Splanwa-Neyman J. On the application of probability theory to agricultural experiments. Essay on Principles. Section 9 (1923). *Statist Sci* 1990;**5**:465–80.
46. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized treatments. *J Educ Psychol* 1974;**66**:688–701.
47. Copas J. Randomization models for the matched and unmatched 2×2 tables. *Biometrika* 1973;**60**:467–76.
48. Miettinen OS. *Theoretical epidemiology*. New York: Wiley, 1985:251–2.
49. Robins JM, Morgenstern H. The foundations of confounding in epidemiology. *Comput Math Applic* 1987;**14**:869–916.
50. Holland P. Statistics and causal inference (with discussion). *J Am Med Assoc* 1986;**81**:940–70.
51. Greenland S. Randomization, statistics, and causal inference. *Epidemiology* 1990;**1**:421–9.
52. Rubin DB. [On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9.] Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies. *Statist Sci* 1990;**5**:472–80.
53. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;**15**:615–25.
54. Greenland S, Robins JM. Identifiability, exchangeability and confounding revisited. *Epidemiologic Perspectives & Innovations* 2009;**6**:4.