

Epidemiology series

Cohort studies: marching towards outcomes

David A Grimes, Kenneth F Schulz

A cohort study tracks two or more groups forward from exposure to outcome. This type of study can be done by going ahead in time from the present (prospective cohort study) or, alternatively, by going back in time to comprise the cohorts and following them up to the present (retrospective cohort study). A cohort study is the best way to identify incidence and natural history of a disease, and can be used to examine multiple outcomes after a single exposure. However, this type of study is less useful for examination of rare events or those that take a long time to develop. A cohort study should provide specific definitions of exposures and outcomes: determination of both should be as objective as possible. The control group (unexposed) should be similar in all important respects to the exposed, with the exception of not having the exposure. Observational studies, however, rarely achieve such a degree of similarity, so investigators need to measure and control for confounding factors. Reduction of loss to follow-up over time is a challenge, since differential losses to follow-up introduce bias. Variations on the cohort theme include the before-after study and nested case-control study (within a cohort study). Strengths of a cohort study include the ability to calculate incidence rates, relative risks, and 95% CIs. This format is the preferred way of presenting study results, rather than with *p* values.

The term cohort has military, not medical, roots. A cohort was a 300–600-man unit in the Roman army; ten cohorts formed a legion (figure 1). The etymology of the term provides a useful mnemonic: a cohort study consists of bands or groups of persons marching forward in time from an exposure to one or more outcomes.

This analogy might be helpful, since cohort studies have a bevy of confusing synonyms: incidence, longitudinal, forward-looking, follow-up, concurrent, and prospective study.^{1,2} Although the terminology can seem daunting, the cohort study is easy for clinicians to understand, since it flows in a logical direction (unlike the case-control study). Here, we explain the terminology, describe the strengths and weaknesses of cohort studies, consider several logistical concerns, mention two permutations of cohort studies, and summarise their analysis.

Data collection: forwards and backwards

A cohort study follows-up two or more groups from exposure to outcome. In its simplest form, a cohort study compares the experience of a group exposed to some factor with another group not exposed to the factor. If the former group has a higher or lower frequency of an outcome than the unexposed, then an association between exposure and outcome is evident.

The defining characteristic of all cohort studies is that they track people forward in time from exposure to outcome. Researchers doing this kind of study must, therefore, go forward in time from the present or go back in time to choose their cohorts (figure 2). Either way, a cohort study moves in the same direction, although gathering data might not. For example, an investigator who wants to study the epidemic of multiple births stemming from assisted reproductive technologies³ could begin a cohort study now. Women exposed to these technologies and a similar group who conceived naturally



Figure 1: An early cohort in search of favourable outcomes

could be tracked forward through their pregnancies to monitor the frequency of multiple births (a concurrent cohort study). Alternatively, the investigator might use existing medical records and go back in time several years to identify women exposed and not exposed to these technologies. He would then track them forward through records to note the birth outcomes. Again, the study moves from exposure to outcome, though the data collection occurred after the fact.

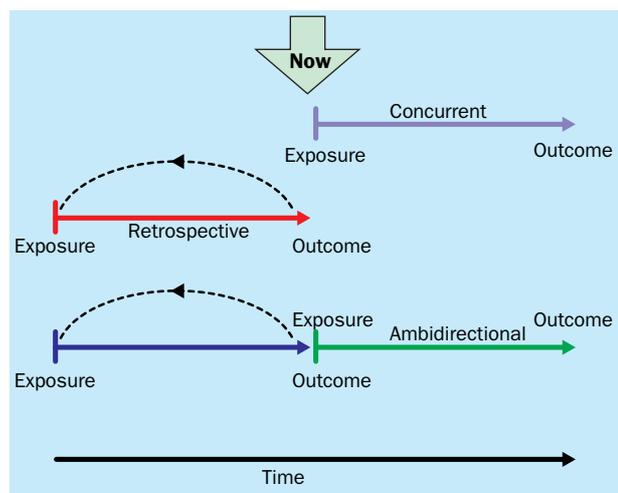


Figure 2: Schematic diagram of concurrent, retrospective, and ambidirectional cohort studies

Lancet 2002; **359**: 341–45

Family Health International, PO Box 13950, Research Triangle Park, NC 27709, USA (D A Grimes MD, K F Schulz PhD)

Correspondence to: Dr David A Grimes (e-mail: dgrimes@fhi.org)

Yet a third variation exists: ambidirectional.⁴ As the name implies, data collection goes in both directions. This approach can be useful for exposures that have both short-term and long-term outcomes. In this hypothetical example, assisted reproductive technologies might be associated with multiple births and with ovarian cancer in later life.⁵ The investigator might, therefore, look back through records for multiple births and also start to follow-up these women into the future for ovarian cancer occurrence.

Advantages of cohort studies

Cohort studies have many appealing features. They are the best way to ascertain both the incidence and natural history of a disorder.⁶ The temporal sequence between putative cause and outcome is usually clear: the exposed and unexposed can often be seen to be free of the outcome at the outset. By contrast, this chicken-egg question often frustrates cross-sectional and case-control studies. For example, in a case-control study, patients with chronic widespread pain were more likely to have mental illness than controls.⁷ Do mood and anxiety disorders increase this risk, or do patients with chronic pain develop mood and anxiety disorders as a result of their disorder?

Cohort studies are useful in investigation of multiple outcomes that might arise after a single exposure. A prototype would be cigarette smoking (the exposure) and stroke, emphysema, oral cancer, and heart disease (the outcomes). Although assessment of many outcomes is often cited as a positive attribute of cohort studies, this feature can be abused. For example, testing the associations between exposure and many outcomes, but only reporting the significant ones, represents misleading science. Investigators should preferably have planned primary and secondary associations to examine (sometimes called hypothesis confirmation). Although investigators can look at other outcomes (hypothesis generation), they should report the findings of all examinations, not just significant ones, so that readers can correctly interpret the results.

The cohort design is also useful in the study of rare exposures: a researcher can often recruit people with uncommon exposures—eg, to ionising radiation or chemicals—in the workplace. A hospital or factory might provide a large number of individuals with the exposure of interest, which would be rare in the general population. Since the investigator does not assign exposure, no ethical concerns arise.

Cohort studies also reduce the risk of survivor bias.⁶ Diseases that are rapidly fatal are difficult to study because of this factor. For example, a hospital-based case-control study of the link between snow-shovelling and myocardial infarction would miss all those who died in the driveway. A cohort study would be a less biased (but more cumbersome) approach: compare rates of myocardial infarction among those who shovel and those who do not shovel. Finally, cohort studies allow calculation of incidence rates, relative risks, and confidence intervals.² Other outcome measures in cohort studies include life-table rates, survival curves, and hazard ratios (panel 1).^{8–10} By contrast, case-control studies cannot provide incidence rates; at best, odds ratios approximate relative risks only when the outcome is uncommon.

Disadvantages of cohort studies

Cohort studies have important limitations too. Selection bias is built into cohort studies. For example, in a cohort study investigating effects of jogging on cardiovascular

Panel 1: Reporting time-to-event in cohort studies

Survival analysis

Survival analysis is useful when lengths of follow-up vary substantially or when participants enter a study at different times.⁸ The Kaplan-Meier method provides a more sophisticated expression of the risk of the outcome over time than does a simple dichotomous outcome.⁹ It can determine the probability (P) of the outcome at any point in time; this result is graphed as a step function (which jumps at every event). A complementary, mirror-image graph portrays the likelihood of avoiding the outcome (1–P) as a function of time (Kaplan-Meier survival curve). The log-rank test compares survival curves of different groups.¹⁰

Proportional hazard model

Another approach to different lengths of follow-up is the Cox proportional hazard model. It is a multivariate technique that has time-to-event (such as illness) as the dependent variable. By contrast, multiple logistic regression has "yes-no" as the dependent variable.⁸ Coefficients from this model can be used to calculate the risk ratio (hazard ratio) of the outcome, after controlling for other covariates in the equation. The hazard ratio (with 95% CIs) is interpreted in the same way as a relative risk for dichotomous outcomes.¹⁰

disease, those who choose to jog probably differ in other important ways (such as diet and smoking) from those who do not exercise.¹¹ In theory, both groups should be the same in all important respects, except for the exposure of interest (jogging), but this seldom occurs. The cohort design is not optimum for rare diseases—eg, scleroderma—or those that take a long time to develop—eg, cancer. However, several large (and thus expensive) cohort studies have made landmark contributions to our knowledge of uncommon diseases. Examples include the Royal College of General Practitioners' Oral Contraceptive Study,¹² the Framingham Heart Study,¹³ the Nurses Health Study,¹⁴ and the British Physicians' Study.¹⁵

Loss to follow-up can be a difficulty, even at 1 month, and particularly so with longitudinal studies that continue for decades. Differential losses to follow-up between those exposed and unexposed can bias results. Over time, the exposure status of study participants can change. For example, a proportion of women who use oral contraceptives will switch to an intrauterine device, and vice versa.¹⁶ Partitioning might be needed to avoid a blurring of exposure, sometimes termed contamination.

What to look for in cohort studies

Who is at risk?

All participants (both exposed and unexposed) in a cohort study must be at risk of developing the outcome.⁶ For example, since women who have had a tubal sterilisation operation have almost no risk of salpingitis,¹⁷ they should not be included in cohort studies of pelvic inflammatory disease.

Who is exposed?

Cohort studies need a clear, unambiguous definition of the exposure at the outset. This definition sometimes involves quantifying the exposure by degree, rather than just yes or no. For example, the minimum exposure might have to be 14 cigarettes per day or less,¹⁸ or 3–6 months of oral contraceptives.¹⁹ Definition of exposure levels in this way can result in more than two groups—eg, non-smokers, light smokers, and heavy smokers.¹⁸

Who is an appropriate control?

The key notion is that controls (the unexposed) should be similar to the exposed in all important respects, except for the lack of exposure. If so, the unexposed group will reveal the background rate of the outcome in the community.

The unexposed group can come from either internal (persons from the same time and place, such as a hospital ward) or external sources. Internal comparisons are most desirable. In a particular population, individuals segregate by themselves (or through medical interventions) into exposure status—eg, cigarette smoking, occupation, contraception. For example, in a cohort study, 138 patients with HIV-1-associated Kaposi's sarcoma were divided into two groups: those with oral and those with cutaneous lesions. The presence of oral lesions (the exposure) had a poorer prognosis, with a median survival (the outcome) one-third that of the other group.²⁰

If satisfactory internal controls are not available, researchers look elsewhere (sometimes termed a double-cohort study).⁶ In a trial of an occupational exposure, finding an adequate number of employees in the factory without the exposure might be difficult. Hence, one might choose workers in a similar factory in the same community. This choice assumes that workers in the other factory have the same baseline risk of the outcome in question, which might not be the case. Even less desirable is use of population norms; disease-specific mortality rates are an example. A researcher might compare lung-cancer death rates among workers in the factory with rates of persons of the same age and sex in the population. Bias inevitably creeps into such comparisons because of the healthy worker effect: those who work are healthier, in general, than those who do not (or cannot) work.^{4,9} Additionally, work reaps economic benefits which might further bias comparisons.

Have outcomes been assessed equally?

Outcomes must be defined in advance; they should be clear, specific, and measurable. Identification of outcomes should be comparable in every way for the exposed and unexposed to avoid information bias. Failure to define objective outcomes leads to uninterpretable results. This challenge relates not only to subjective syndromes such as Gulf War,²¹ chronic fatigue,^{22,23} and premenstrual,²⁴ but also to more mundane health problems such as endometritis. Just how tender must a uterus be? Keeping those who judge outcomes unaware of the exposure status of participants (blinding) in a cohort study is important for subjective outcomes, such as tenderness or erythema. By contrast, with objective outcome measures, such as fever or death, blinding the exposure status is less important.

Outcome information can come from many sources. For mortality studies, the death certificate is often used. Although convenient, the validity of the clinical information is highly variable. For non-fatal outcomes, sources include hospital charts, insurance records, laboratory records, disease registries, hospital discharge logs, and physical examination and measurement of participants. Optimally, the person who judges outcomes should be unaware of the exposure. When diagnoses vary in their confidence, assignment of levels of assurance might be helpful, such as definite, probable, and suspect.⁹

Tracking participants over time*Have losses been minimised?*

Although loss of participants damages the power and precision of a study, differential loss to follow-up is more sinister. Bail-outs are not random events. If the likelihood

of bailing out is related both to exposure and outcome, then bias can result.²⁵ For example, some participants given a new antibiotic might have such poor outcomes that they are unable to complete questionnaires or to return for examination.²⁶ Their disappearance from the cohort would make the new antibiotic look better than it is.

The best way of dealing with loss to follow-up is to avoid it. For example, restrict participation to only those judged likely to complete the study. Additionally, several safeguards are customary. Obtaining the names of several family members or friends who do not live with the respondent is often helpful at the start of such studies. The participant's family doctor might also be helpful. Should the respondent move, these contacts would probably know their new address. Motor vehicle registration records can be useful too. Furthermore, national vital statistics registries, such as the National Death Index in the USA, facilitate follow-up. Participants can be offered financial compensation for their time lost from work as a result of the study. Diligent tracking of participants is hard work, and might require hiring personnel for this task alone.

Reporting cohort studies

Many researchers who do cohort studies report their findings in an unsatisfactory way (panel 2).²⁷ An investigator's first challenge is to convince the editor (then readers) that the exposed and unexposed groups were indeed similar in all important respects, except for the exposure. The first table in reports of cohort studies customarily provides demographic and other prognostic factors for both groups with hypothesis testing (p values) to show the likelihood that observed differences could be due to chance.

For dichotomous outcome measures, such as sick or well, the investigator should provide raw data sufficient for the reader to confirm the results. For cumulative incidence, the investigator should calculate the proportion who developed the outcome during the specified study interval. For incidence rates, the value is expressed per unit of time.⁴ Then, relative risks and confidence intervals should be provided. Use of p values should not replace interval estimation (relative risks with confidence

Panel 2: Features to look for in a cohort study**How much selection bias was present?**

- 1 Were only people at risk of the outcome included?
- 1 Was the exposure clear, specific, and measurable?
- 1 Were the exposed and unexposed groups similar in all important respects except for the exposure?

What steps were taken to minimise information bias?

- 1 Was the outcome clear, specific, and measurable?
- 1 Was the outcome identified in the same way for both groups?
- 1 Was determination of outcome made by an observer blinded as to treatment?

How complete was the follow-up of both groups?

- 1 What efforts were made to limit loss to follow-up?
- 1 Was loss to follow-up similar in both groups?

Were potential confounding factors sought and controlled for in the analysis?

- 1 Did the investigators anticipate and gather information on potential confounding factors?
- 1 What method(s) were used to assess and control for confounding?

intervals)²⁸ and should only be used as supplemental information.²⁵

Like other observational studies, cohort studies have built-in bias. Investigators should identify potential biases in their data and show how these might have affected results. Whenever possible, confounding should be controlled for in the analysis. These techniques are discussed in an earlier essay in this series.²⁹

Variations on the cohort theme

Before-after studies

Before-after studies (time series) have important limitations. Here, an investigator takes a measurement, exposes participants to an intervention (often a drug), repeats the measurements, then compares them. First, regression to the mean is often ignored. If admission to the cohort includes extreme measurements,³⁰ such as high laboratory values, then lower mean values will arise at follow-up, irrespective of treatment.³¹ Second, secular trends, such as seasonal changes in the frequency of pneumonia, can affect results. Third, washout periods are often needed to avoid a carryover effect of drugs given during the initial observation period.⁶

Nested case-control studies

Cohort studies sometimes spawn other studies. One of the most frequent is the nested case-control study.^{6,9,25} Why would an investigator carve out a case-control study in the midst of a cohort study? The answer often involves body fluids and a freezer. Some exposure or predictor variables are simply too expensive to determine on everyone in a study. A sophisticated blood test is the prototype. A clever way to skirt this financial obstacle is to do a cohort study that will yield a sufficient number of cases. All participants entering the cohort study have a tube of blood drawn at enrolment; serum is frozen until the study's conclusion. All those in the cohort study who develop the outcome of interest now become the cases for the nested study. The investigator then chooses a random sample of all participants who did not develop the outcome (controls). Next, the blood test is done on serum from only the cases and controls, not the whole group of exposed and unexposed. In this way, the laboratory cost is minimised while assuring that the exposure—eg, a positive laboratory test—was present before development of the outcome. Controls are generally matched to cases by important characteristics, such as age and sex.⁹

A nested case-control study, for example, examined the potential relation between body concentrations of organochlorines and non-Hodgkin's lymphoma. The blood samples were obtained on entry to a large cohort study started in Maryland, USA, in 1974. Blood samples were eventually analysed for only 74 individuals with lymphoma and 147 controls.³² Thus, instead of measuring organochlorine concentrations of the entire cohort of 25 802, the investigators incurred this laboratory expense for less than 1% of the cohort. In view of the availability of banked blood specimens around the world, this type of research design is likely to become popular. However, nested case-control studies might be useful for other studies that do not require blood tests but in which determination of the exposure is expensive or difficult⁹—eg, measurement of nerve conduction³³ or job stressors.³⁴

Conclusion

Cohort studies are common in medical research. Like other research designs, they entail important trade-offs. Readers should make sure that investigators provide clear,

specific, and measurable definitions of exposures and outcomes. The unexposed group should resemble the exposed group in all important respects, and determination of outcomes should be objective and, whenever possible, blinded. Results for dichotomous outcomes should be provided as rates, relative risks, and confidence intervals, which offer more information than do p values. Reports of cohort studies should identify and describe the potential effect of biases. Importantly, investigators should measure and control for potential confounding.

We thank Willard Cates and David L Sackett for their helpful comments on an earlier version of this report. Much of this material stems from our 15 years of teaching the Berlex Foundation Faculty Development Course.

References

- 1 Last JM, ed. A dictionary of epidemiology, 2nd edn. New York: Oxford University Press, 1988.
- 2 Lilienfeld AM, Lilienfeld DE. Foundations of epidemiology, 2nd edn. New York: Oxford University Press, 1980.
- 3 Anon. Contribution of assisted reproductive technology and ovulation-inducing drugs to triplet and higher-order multiple births—United States, 1980–1997. *MMWR Morb Mortal Wkly Rep* 2000; **49**: S35–38.
- 4 Hennekens CH, Buring JE. Epidemiology in medicine. Boston: Little, Brown and Company, 1987.
- 5 Whittemore AS, Harris R, Itnyre J. Characteristics relating to ovarian cancer risk: collaborative analysis of 12 US case-control studies, II: invasive epithelial ovarian cancers in white women. Collaborative Ovarian Cancer Group. *Am J Epidemiol* 1992; **136**: 1184–203.
- 6 Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman RB, eds. Designing clinical research: an epidemiologic approach, 2nd edn. Baltimore: Lippincott Williams and Wilkins, 2001.
- 7 Benjamin S, Morris S, McBeth J, Macfarlane GJ, Silman AJ. The association between chronic widespread pain and mental disorder: a population-based study. *Arthritis Rheum* 2000; **43**: 561–67.
- 8 Wassertheil-Smoller S. Biostatistics and epidemiology: a primer for health professionals, 2nd edn. New York: Springer-Verlag, 1995.
- 9 Kelsey JL, Whittemore AS, Evans AS, Thompson WD. Methods in observational epidemiology, 2nd edn. New York: Oxford University Press, 1996.
- 10 Lang TA, Secic M. How to report statistics in medicine. Philadelphia: American College of Physicians, 1997.
- 11 Sackett DL. Bias in analytic research. *J Chron Dis* 1979; **32**: 51–63.
- 12 Beral V, Hermon C, Kay C, Hannaford P, Darby S, Reeves G. Mortality associated with oral contraceptive use: 25 year follow up of cohort of 46 000 women from Royal College of General Practitioners' oral contraception study. *BMJ* 1999; **318**: 96–100.
- 13 Seman LJ, DeLuca C, Jenner JL, et al. Lipoprotein(a)-cholesterol and coronary heart disease in the Framingham Heart Study. *Clin Chem* 1999; **45**: 1039–46.
- 14 Colditz GA, Rosner BA, Speizer FE. Risk factors for breast cancer according to family history of breast cancer. *J Natl Cancer Inst* 1996; **88**: 365–71.
- 15 Doll R, Peto R, Boreham J, Sutherland I. Smoking and dementia in male British doctors: prospective study. *BMJ* 2000; **320**: 1097–102.
- 16 Hannaford PC, Kay CR. The risk of serious illness among oral contraceptive users: evidence from the RCGP's oral contraceptive study. *Br J Gen Pract* 1998; **48**: 1657–62.
- 17 Levgur M, Duvivier R. Pelvic inflammatory disease after tubal sterilization: a review. *Obstet Gynecol Surv* 2000; **55**: 41–50.
- 18 Croft P, Hannaford P. Risk factors for acute myocardial infarction in women. *BMJ* 1989; **298**: 674.
- 19 Anon. The reduction in risk of ovarian cancer associated with oral-contraceptive use. The Cancer and Steroid Hormone Study of the Centers for Disease Control and the National Institute of Child Health and Human Development. *N Engl J Med* 1987; **316**: 650–55.
- 20 Rohrmus B, Thoma-Greber EM, Bogner JR, Rocken M. Outlook in oral and cutaneous Kaposi's sarcoma. *Lancet* 2000; **356**: 2160.
- 21 Fukuda K, Nisenbaum R, Stewart G, et al. Chronic multisymptom illness affecting Air Force veterans of the Gulf War. *JAMA* 1998; **280**: 981–98.
- 22 Schluenderberg A, Straus SE, Peterson P, et al. NIH conference. Chronic fatigue syndrome research: definition and medical outcome assessment. *Ann Intern Med* 1992; **117**: 325–31.
- 23 Armon C, Kurland LT. Chronic fatigue syndrome: issues in the diagnosis and estimation of incidence. *Rev Infect Dis* 1991; **13** (suppl 1): S68–72.

- 24 Haskett RF, Abplanalp JM. Premenstrual tension syndrome: diagnostic criteria and selection of research subjects. *Psychiatr Res* 1983; **9**: 125–38.
- 25 Rothman KJ. *Modern epidemiology*. Boston: Little, Brown and Company, 1986.
- 26 Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical epidemiology: a basic science for clinical medicine*, 2nd edn. Boston: Little, Brown and Company, 1991.
- 27 Bracken MB. Reporting observational studies. *Br J Obstet Gynaecol* 1989; **96**: 383–88.
- 28 Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. *BMJ* 1986; **292**: 746–50.
- 29 Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet* 2002; **359**: 248–52.
- 30 Ewan PW, Clark AT. Long-term prospective observational study of patients with peanut and nut allergy after participation in a management plan. *Lancet* 2001; **357**: 111–15.
- 31 Bland JM, Altman DG. Regression towards the mean. *BMJ* 1994; **308**: 1499.
- 32 Rothman N, Cantor KP, Blair A, et al. A nested case-control study of non-Hodgkin lymphoma and serum organochlorine residues. *Lancet* 1997; **350**: 240–44.
- 33 Atterbury MR, Limke JC, Lemasters GK, et al. Nested case-control study of hand and wrist work-related musculoskeletal disorders in carpenters. *Am J Ind Med* 1996; **30**: 695–701.
- 34 Kawakami N, Araki S, Kawashima M. Effects of job stress on occurrence of major depression in Japanese industry: a case-control study nested in a cohort study. *J Occup Med* 1990; **32**: 722–25.

Uses of error

The error cascade

Neil Gittoes

Having just been appointed consultant physician, I found myself reflecting on my career and realising that I could at last stand alone and that finally the buck stops with me. There was a time that I wished it didn't. I was a medical senior house officer when I saw an elderly man who described a subacute onset of breathlessness and a dry cough. He had trouble speaking and was using his accessory muscles. He had initially received standard nebulised treatment for exacerbated chronic obstructive pulmonary disease, although his chest radiograph showed a small pneumothorax on the left. After conferring with senior colleagues, I inserted a chest drain on the left, and verified its position with a second radiograph. In the middle of the night the house officer saw the patient with worsening shortness of breath and surgical emphysema. He pushed the tube in further, but an hour later the arrest team were called because the patient had developed extreme respiratory distress and had become cyanosed. They thought that he had developed a contralateral

pneumothorax and proceeded to insert a chest drain on the right. Arriving on the ward the following morning, I was horrified to find my patient with bilateral chest drains and surgical emphysema from head to scrotum. However, at least he was alive. Chest radiographs and computed tomography showed bilateral pneumothoraces with both drains embedded deeply within the lung parenchyma, just short of the mediastinum on the right, and abutting the left ventricle on the left. I inserted bilateral anterior drains and cautiously removed the lateral ones. After a few days the right-sided pneumothorax resolved, although the left side needed surgical correction. He was finally discharged, and on reviewing the radiographs it was apparent that I had inserted the original drain where there was a small area of pleural adhesion. The two pleural surfaces remained contiguous, and the drain entered the lung parenchyma. The subsequent errors of management turned the situation rapidly into a life-threatening predicament.

Queen Elizabeth Hospital, Edgbaston, Birmingham, B15 2 TH UK (N Gittoes MRCP)