

4. **Long FR**, Williams RS, Castile RG. Structural airway abnormalities in infants and young children with cystic fibrosis. *J Pediatr* 2004;**144**:154–61.
5. **Linnane BM**, Hall GL, Nolan G, *et al*; on behalf of the Australian Respiratory Early Surveillance Team for Cystic Fibrosis (AREST-CF). Lung function in infants with cystic fibrosis diagnosed by newborn screening. *Am J Respir Crit Care Med* 2008;**178**:1238–44.
6. **Sly PD**, Brennan S, Gangel C, *et al*; on behalf of the Australian Respiratory Early Surveillance Team for Cystic Fibrosis (AREST-CF). Lung disease at diagnosis in infants with cystic fibrosis detected by newborn screening. *Am J Respir Crit Care Med* 2009;**180**:146–52.
7. **Armstrong DS**, Grimwood K, Carlin JB, *et al*. Lower airway inflammation in infants and young children with cystic fibrosis. *Am J Respir Crit Care Med* 1997;**156**:1197–204.
8. **Dakin CJ**, Numa AH, Wang H, *et al*. Inflammation, infection, and pulmonary function in infants and young children with cystic fibrosis. *Am J Respir Crit Care Med* 2002;**165**:904–10.
9. **Sagel SD**, Gibson RL, Emerson J, *et al*; For the Inhaled Tobramycin in Young Children Study Group and the Cystic Fibrosis Foundation Therapeutics Development Network. Impact of Pseudomonas and Staphylococcus infection on inflammation and clinical status in young children with cystic fibrosis. *J Pediatr* 2009;**154**:183–8.
10. **Chmiel JF**, Berger M, Konstan MW. The role of inflammation in the pathophysiology of CF lung disease. *Clin Rev Allergy Immunol* 2002;**23**:5–27.
11. **Armstrong DS**, Hook SM, Jansen KM, *et al*. Lower airway inflammation in infants with cystic fibrosis detected by newborn screening. *Pediatr Pulmonol* 2005;**40**:500–10.
12. **Douglas TA**, Brennan S, Gard S, *et al*. Acquisition and eradication of *P. aeruginosa* in young children with cystic fibrosis. *Eur Respir J* 2009;**33**:305–11.
13. **Rosenfeld M**, Emerson J, Accurso F, *et al*. Diagnostic accuracy of oropharyngeal cultures in infants and young children with cystic fibrosis. *Pediatr Pulmonol* 1999;**28**:321–8.
14. **Kosorok MR**, Zeng L, West SEH, *et al*. Acceleration of lung disease in children with cystic fibrosis after Pseudomonas aeruginosa acquisition. *Pediatr Pulmonol* 2001;**32**:277–87.
15. **Nixon GM**, Armstrong DS, Carzino R, *et al*. Clinical outcome after early Pseudomonas aeruginosa infection in cystic fibrosis. *J Pediatr* 2001;**138**:699–704.
16. **Emerson J**, Rosenfeld M, McNamara S, *et al*. Pseudomonas aeruginosa and other predictors of mortality and morbidity in young children with cystic fibrosis. *Pediatr Pulmonol* 2002;**34**:91–100.
17. **Li Z**, Kosorok MR, Farrell PM, *et al*. Longitudinal development of mucoid Pseudomonas aeruginosa infection and lung disease progression in children with cystic fibrosis. *JAMA* 2005;**293**:581–8.
18. **Courtney JM**, Bradley J, McCaughan J, *et al*. Predictors of mortality in adults with cystic fibrosis. *Pediatr Pulmonol* 2007;**42**:525–32.
19. **Johansen HK**, Göttsche PC. Vaccines for preventing infection with Pseudomonas aeruginosa in cystic fibrosis. *Cochrane Database Syst Rev* 2008;(4):CD001399.
20. **Döring G**, Høiby N; For the Consensus Study Group. Early intervention and prevention of lung disease in cystic fibrosis: a European consensus. *J Cyst Fibros* 2004;**3**:67–91.
21. **Heinzel B**, Eber E, Oberwaldner B, *et al*. Effects of inhaled gentamicin prophylaxis on acquisition of Pseudomonas aeruginosa in children with cystic fibrosis: a pilot study. *Pediatr Pulmonol* 2002;**33**:32–7.
22. **Lebecque P**, Leal T, Zylberberg K, *et al*. Towards zero prevalence of chronic Pseudomonas aeruginosa infection in children with cystic fibrosis. *J Cyst Fibros* 2006;**5**:237–44.
23. **Tramper-Stranders GA**, Wolfs TFW, van Haren Noman S, *et al*. Controlled trial of cycled antibiotic prophylaxis to prevent initial Pseudomonas aeruginosa infection in children with cystic fibrosis. *Thorax* 2010;**65**:897–902.
24. **Valerius N**, Koch C, Høiby N. Prevention of chronic Pseudomonas aeruginosa colonisation in cystic fibrosis by early treatment. *Lancet* 1991;**338**:725–6.
25. **Armstrong DS**, Grimwood K, Carlin JB, *et al*. Bronchoalveolar lavage or oropharyngeal cultures to identify lower respiratory pathogens in infants with cystic fibrosis. *Pediatr Pulmonol* 1996;**21**:267–75.
26. **Gutierrez JP**, Grimwood K, Armstrong DS, *et al*. Interlobar differences in bronchoalveolar lavage fluid from children with cystic fibrosis. *Eur Respir J* 2001;**17**:281–6.
27. **Ratjen F**, Walter H, Haug M, *et al*. Diagnostic value of serum antibodies in early Pseudomonas aeruginosa infection in cystic fibrosis patients. *Pediatr Pulmonol* 2007;**42**:249–55.
28. **Langton Hewer SC**, Smyth AR. Antibiotic strategies for eradicating Pseudomonas aeruginosa in people with cystic fibrosis. *Cochrane Database Syst Rev* 2009;(4):CD004197.
29. **Jones AM**. Eradication therapy for early Pseudomonas aeruginosa infection in CF: many questions still unanswered. *Eur Respir J* 2005;**26**:373–5.
30. **Eber E**, Thalhammer GH, Zach MS. Eradication of Pseudomonas aeruginosa in cystic fibrosis. *Eur Respir J* 2006;**27**:438–9.
31. **Gibson RL**, Emerson J, Mayer-Hamblett N, *et al*; For the Inhaled Tobramycin in Young Children Study Group and the Cystic Fibrosis Foundation Therapeutics Development Network. Duration of treatment effect after tobramycin solution for inhalation in young children with cystic fibrosis. *Pediatr Pulmonol* 2007;**42**:610–23.
32. **Ratjen F**, Munck A, Kho P, *et al*; For the ELITE Study Group. Treatment of early Pseudomonas aeruginosa infection in patients with cystic fibrosis: the ELITE trial. *Thorax* 2010;**65**:286–91.
33. **Que C**, Cullinan P, Geddes D. Improving rate of decline of FEV1 in young adults with cystic fibrosis. *Thorax* 2006;**61**:155–7.
34. **Konstan MW**, Morgan WJ, Butler SM, *et al*; For the Scientific Advisory Group and the Investigators and Coordinators of the Epidemiologic Study of Cystic Fibrosis. Risk factors for rate of decline in forced expiratory volume in one second in children and adolescents with cystic fibrosis. *J Pediatr* 2007;**151**:134–9.
35. **Daviskas E**, Anderson SD, Brannan JD, *et al*. Inhalation of dry-powder mannitol increases mucociliary clearance. *Eur Respir J* 1997;**10**:2449–54.
36. **Jaques A**, Daviskas E, Turton JA, *et al*. Inhaled mannitol improves lung function in cystic fibrosis. *Chest* 2008;**133**:1388–96.
37. **Daviskas E**, Anderson SD, Jaques A, *et al*. Inhaled mannitol improves the hydration and surface properties of sputum in patients with cystic fibrosis. *Chest* 2010;**137**:861–8.
38. **Minasian C**, Wallis C, Metcalfe C, *et al*. Comparison of inhaled mannitol, daily rDNase and a combination of both in children with cystic fibrosis: a randomised trial. *Thorax* 2010;**65**:51–6.
39. **Frederikson B**, Pressler T, Hansen A, *et al*. Effect of aerosolized rDNase (Pulmozyme®) on pulmonary colonization in patients with cystic fibrosis. *Acta Paediatr* 2006;**95**:1070–4.
40. **Ring E**, Eber E, Erwa W, *et al*. Urinary N-acetyl-β-D-glucosaminidase activity in patients with cystic fibrosis on long term gentamicin inhalation. *Arch Dis Child* 1998;**78**:540–3.
41. **Bockenbauer D**, Hug MJ, Kleta R. Cystic fibrosis, aminoglycoside treatment and acute renal failure: the not so gentle micin. *Pediatr Nephrol* 2009;**24**:925–8.
42. **Emerson J**, McNamara S, Buccat AM, *et al*. Changes in cystic fibrosis sputum microbiology in the United States between 1995 and 2008. *Pediatr Pulmonol* 2010;**45**:363–70.
43. **Treggiari MM**, Rosenfeld M, Mayer-Hamblett N, *et al*; EPIC Study Group. Early anti-pseudomonal acquisition in young patients with cystic fibrosis: rationale and design of the EPIC clinical trial and observational study. *Contemp Clin Trials* 2009;**30**:256–68.

Relearning an old lesson: stopping trials early

Najib M Rahman, Robert J O Davies

UKCRC Oxford Respiratory Trials Unit, University of Oxford and Oxford Centre for Respiratory Medicine, Oxford Radcliffe Hospital, Oxford, UK

Correspondence to Robert J O Davies, Oxford Radcliffe Hospital and Oxford University, Churchill Hospital, Headington, Oxford OX3 7LJ, UK; robert.davies@ndm.ox.ac.uk

A well designed and delivered clinical trial is the main tool to define whether medical interventions ‘work’, and how well. As such, they are potent weapons in the armoury of medical progress—and like all potent weapons need to be used with care.

In this month’s Thorax, Koegelenberg *et al* (*see page 857*) report the findings of a trial comparing the diagnostic accuracy of closed pleural biopsy (Abrams needle) and cutting needle pleural biopsy after thoracic ultrasound, for the diagnosis of pleural tuberculosis (TB).¹ This question is clearly important given the global significance of TB, and the key role of pleural biopsy in the diagnosis and microbiological assessment of its pleural presentations. To date, there are no published studies assessing ultrasound-guided pleural biopsy for the diagnosis of

TB-related pleural effusions. This study is a continuation of this group's research programme which has a track record of delivering valuable evidence in the diagnosis of pleural TB, not least their previous study showing that thoracoscopy is superior to closed pleural biopsy in this disease.²

Their studies are conducted in an area with a high prevalence of TB, with all the recruited subjects receiving the compared diagnostic tests, allowing the comparison of diagnostic results. In this new study, all patients underwent both Abrams biopsy and cutting needle biopsy, performed in random order. The design is simple, logical and efficient, with the diagnostic accuracy for TB assessed against accepted 'gold standards'. Given the biology of pleural TB, which manifests as diffuse pleural involvement, it is reasonable to propose that the two biopsy techniques may be similar. As such, the results of the study should be clinically important, but on this occasion caution needs to be exercised in interpreting these results in view of a flaw in study delivery.

The study was stopped prematurely (after 89 (40%) of a planned 220 subject recruitment). The study was originally, appropriately, planned as a non-inferiority design requiring 220 patients to demonstrate 'equivalence' of the two techniques within a 10% threshold. It was halted when a statistically significant difference was identified at a preplanned interim analysis, with closed pleural biopsy appearing more sensitive than cutting needle biopsy (81.8% vs 65.2%, $p=0.022$). Unfortunately, this early cessation means it is difficult to interpret the result of the study and assess the possible benefit in favour of closed biopsy.

A clinical trial would normally only be stopped early in the face of 'overwhelming evidence' of a difference in outcome between the study groups (conventionally a p value of about <0.0001), a criterion generally known as the Peto–Haybittle rule.^{3 4} The stopping decision would normally be taken by an independent group who are not part of the core investigator team, to avoid bias. This is key to the ability of the trial to fulfil its primary functions of delivering a result that is highly likely to be true, and is useful in quantifying the magnitude of any benefits seen. It is worth revisiting why this is so.

There are several reasons why trials should not stop when a statistically significant difference is first seen. 'Statistically significant' differences commonly arise by chance early during trial recruit-

ment, and disappear later. This is the reason for the demanding ($p<0.0001$) threshold in the Peto–Haybittle rule. In this study, the statistical signal was $p=0.022$, creating the possibility that the result is a statistical fluke and hence wrong. If the next two cases recruited to the study happened to favour cutting needle biopsy, the statistical significance would have disappeared ($p=0.08$).

Our Unit's trial of adjuvant intrapleural streptokinase in pleural infection⁵ is a real-life example of early and misleading statistical significance. During recruitment to this trial the independent Data Monitoring Committee first reviewed the data for safety assessment reasons after ~40% of the recruitment. At this review there was a 'statistically significant' difference between the study groups in the frequency of death and surgery, with a p value of ~0.01. This was not thought to constitute 'overwhelming evidence' of a treatment difference, was not communicated to the trial team and recruitment was allowed to proceed. By the next review this difference had disappeared and the eventual trial result was completely negative. With the benefit of hindsight, if the study had been stopped and published when this first significant p value was identified, we would have reported a result that was untrue.

Secondly, if a statistically significant difference in the trial outcome between study arms is used as a reason to stop, it is tempting to assess the data repeatedly in order to deliver the study result rapidly. Unfortunately, every 'peek' at the data increases the likelihood of a false-positive result. The conventional maximum p value for (arbitrary) statistical significance is <0.05 , which means that the observed result has a chance of having arisen fortuitously of <1 in 20 (5%). However, if the results are assessed on 20 different occasions, a p value of 0.05 will occur by statistical chance alone on one occasion. The study investigators in this study did preplan their interim analysis, but they did not adjust their p values for this extra 'peek' at the data. A correction for one 'peek' would imply a p value of ~0.025—barely achieved when the trial was stopped.

Finally, stopping trials earlier than planned reduces the precision of the estimate of the outcome being assessed. This is most commonly expressed using the 95% CI for the result. In this study, the observed difference in the diagnostic sensitivity between the two techniques is 16.7%; however, the CIs imply that it is somewhere between 1.9% and 31.5%. This means that there is anything from an

enormous and vital advantage, through to a trivial and unimportant difference in using Abrams biopsy for the diagnosis of TB. If recruitment had been completed to the original target of 220 and the difference between the groups had remained unaltered, the result would have allowed us to be reasonably confident that the advantage to Abrams biopsy would lie between 10.9% and 27.3%—a more clinically useful estimate.

So how should the clinician respond to these results? Despite the above limitations, these are the first high quality randomised data in this area and the authors should be congratulated on conducting such a difficult and intensive trial. The study also provides interesting data on the high diagnostic sensitivity for pleural malignancy using an ultrasound-guided closed pleural biopsy strategy (100% for combined Abrams and Tru-cut)—this aspect now warrants further specific study. The results of this study suggest that there may be an advantage to Abrams biopsy over cutting needle biopsy for the diagnosis of TB—as this disease is highly prevalent in resource-poor areas, this would be an important finding with implications for practice. However, we cannot be confident that this is true or of the magnitude of any possible advantage. Accordingly, clinical practice should not change on the basis of these results. This study demonstrates that such a trial is deliverable, and that there is a significant possibility that a more simple diagnostic strategy may be advantageous, consistent with previous evidence on the relatively high yield of Abrams biopsies in the diagnosis of pleural TB. Given the lack of data in this field, a fully completed randomised study of similar design would be the appropriate next step to advance diagnostic strategies in this important disease, and this study provides accurate data on the sample size required for such a trial. If such a study again demonstrated high diagnostic yields for both TB and malignancy with an ultrasound-guided blind technique, future assessment of this strategy compared with thoracoscopy or CT-guided biopsy would be warranted.

Competing interests None.

Provenance and peer review Commissioned; not externally peer reviewed.

Thorax 2010;65:851–853.
doi:10.1136/thx.2009.131219

REFERENCES

1. Koegeleberg CFN, Bolliger CT, Theron J, *et al.* A direct comparison of the diagnostic yield of ultrasound-assisted Abrams and Tru-cut needle

- biopsies for pleural tuberculosis. *Thorax* 2010;**65**: 857–862.
- Diacon AH**, Van de Wal BW, Wyser C, *et al.* Diagnostic tools in tuberculous pleurisy: a direct comparative study. *Eur Respir J* 2003;**22**:589–91.
 - Haybittle J**. Repeated assessment of results in clinical trials of cancer treatment. *Br J Radiol* 1971;**44**:793–7.
 - Peto R**, Pike M, Armitage P. Design and analysis of randomized clinical trials requiring prolonged observation of each patient: introduction. *Br J Cancer* 1976;**34**:585–612.
 - Maskell NA**, Davies CW, Nunn AJ, *et al.* U.K. Controlled trial of intrapleural streptokinase for pleural infection. *N Engl J Med* 2005;**352**:865–74.

Severity scores for CAP. 'Much workload for the next bias'

Santiago Ewig,¹ Antoni Torres²

Tools for the assessment of severity of patients with community-acquired pneumonia (CAP) have attracted much interest in the recent past. In this issue of *Thorax* (see page 878) two systematic reviews and meta-analyses address the value of such tools including no less than 40¹ and 23 studies.² Despite different inclusion strategies and statistical approaches, both studies report two main and similar conclusions. First, both the most extensively investigated tools (Pneumonia Severity Index (PSI) and CURB-65/CRB-65) have remarkably favourable power to predict mortality. Secondly, whereas PSI is somewhat better in predicting patients at low risk, the reverse is true for CURB-65/CRB-65. However, these differences are of questionable clinical relevance. Thus, both tools can be regarded as equivalent. CRB-65 is the simplest tool and can easily be remembered and also applied in the outpatient setting. These straightforward conclusions are the result of a decade of intensive and successful work to establish clinically useful tools of severity assessment. So far success is impressive; clinicians now can use a very simple tool as an adjunct to clinical judgement, and studies on patients with CAP can rely on a validated tool for severity stratification.³

At this point, it is worthwhile having a look at the perspectives beyond: where do we have to go from here? Actually, there are many issues that still must be

addressed concerning validation in intervention studies and still insufficiently recognised ambiguities inherent to the severity scores.

PSI has been prospectively validated in independent populations as a tool to guide site of treatment decisions, and the use of the PSI was associated with a larger proportion of patients in PSI risk classes I and II who were treated in the outpatient environment without compromising their safety.⁴ No such studies have been performed for CURB-65/CRB-65, and obviously these studies need to be performed.

On the other hand, it is difficult to imagine that such a complex tool as the PSI will ever be implemented in routine clinical practice, particularly in the outpatient setting. CURB-65, and particularly CRB-65, is clearly superior in terms of simplicity, and it is its simplicity that makes it preferable even if it provides slightly inferior predictions. However, the four-variable CRB-65 score is a delicate tool and vulnerable to ambiguities of its components. First, the respiratory rate may be difficult to assess in patients with rapid shallow breathing. In fact, no rules have been established regarding how to assess the respiratory rate precisely. Secondly, pneumonia-associated confusion is difficult if not impossible to distinguish from premorbid mental deficiencies in elderly and disabled patients. Evidently, pre-existing and pneumonia-related confusion cannot be confidently differentiated in every case. As a result, confusion cannot be regarded exclusively as a parameter reflecting only severe sepsis but must also be recognised as a parameter possibly reflecting pre-existing mild central nervous system co-morbidity. The substrate of the parameter 'confusion' remains somewhat vague. Thirdly, if CURB-65 is used, blood urea nitrogen may be a confounder in patients with pre-existing renal insufficiency and in the

elderly. In fact, some have found that CURB-65 works less well in elderly patients. In one recent study of hospitalised elderly patients, the area under the curve (AUC) was significantly higher for the <65-year cohort in comparison with older patients (0.93 vs 0.7).⁵ Others have suggested that oxygenation instead of confusion and blood urea nitrogen might work at least as well.⁶ There is evidence that performance status is an independent predictor for short- and long-term mortality in hospitalised elderly patients.⁷ Accordingly, combining the modified PSI with performance status led to better predictions being obtained.⁸ Fourthly, the threshold of age (65 years) seems arbitrary and there is some evidence that higher cut-offs might work better.⁹ Splitting age into decades might be even more promising. However, the rule would clearly lose much of the main strength of simplicity if stratification by age were to form a part of it.

The simplicity of the CRB-65 rule is not without pitfalls. A score of CRB-65 = 1 already implies an increased risk of death and should prompt consideration of hospitalisation. However, when this score is applied in primary care, hospital referral would have to be considered in all patients just because of an age older than 65 years. Although increasingly age >50 years is associated with increased risk of death, it is clearly impractical and inadequate to hospitalise all patients with CAP only because they are older than 65 years.¹⁰ Thus, it is important to exclude age alone as a criterion when CRB-65 is used as an aid for the decision to hospitalise.

Studies addressing severity assessment of CAP mainly refer to hospitalised patients, and clearly more data are needed for the outpatient setting. Obviously, the available tools do not measure the same things and are imperfect, and there has been insufficient attention so far paid to the reasons behind this. Several studies have incorporated biomarkers to improve mortality predictions, and as things stand today it appears that some of these (in particular C-reactive protein and^{11 12} procalcitonin,¹¹ but also adrenomedullin)^{13–15} should be able to meet expectations.

¹Thoraxzentrum Ruhrgebiet, Kliniken für Pneumologie und Infektiologie, EVK Herne und Augusta-Kranken-Anstalt Bochum, Germany; ²Servei de Pneumologia, Institut Clinic del Tòrax, Hospital Clinic de Barcelona, Facultat de Medicina, Universitat de Barcelona, Idibaps, Ciber de Enfermedades Respiratorias (CIBERES), Spain

Correspondence to Prof. Dr. Santiago Ewig, Thoraxzentrum Ruhrgebiet, Kliniken für Pneumologie und Infektiologie, EVK Herne und Augusta-Kranken-Anstalt Bochum, Bergstrasse 26, D-44791 Bochum, Germany; ewig@augusta-bochum.de