Primary Care Respiratory Journal (2010); **19**(2): 170-179

Primary Care RESPIRATORY JOURNAL
www.thepcrj.org

RESEARCH INTO PRACTICE

Lessons from the major studies in COPD: problems and pitfalls in translating research evidence into practice

***David MG Halpin**[a]

[a] Consultant Physician & Honorary Senior Clinical Lecturer in Respiratory Medicine, Royal Devon & Exeter Hospital, Exeter, UK

Received 21st January 2010; accepted 7th February 2010; online 29th March 2010

Summary

Translating the growing evidence base on COPD management into practice can be challenging and understanding the strengths and weakness of published studies is crucial. Studies should conform to the standards of CONSORT statement; they should be sufficiently powered, participants should be randomised, there should be assignment concealment, and the outcome measures and analyses should be decided in advance.

The interpretation of the results may be affected by age and severity inclusion criteria for the study and the exclusion of patients with co-morbid illnesses. Whether previous medication is continued or stopped can affect the interpretation of the results. Secondary analyses in sub-groups should be viewed with caution unless pre-specified and accommodated in the trial design and power calculations. Real world observational studies may be confounded by non-randomisation of participants but can sometimes yield valuable insights.

The way in which the results are presented can influence their interpretation and their magnitude with respect to minimal important differences as well as statistical significance is important.

Research studies help formulate management algorithms but often the questions they address are too specific to allow evidence-based sequencing of therapies.

DMG Halpin. *Prim Care Resp J* 2010; **19**(2): 170-179
doi:10.4104/pcrj.2010.00015

Keywords COPD, trials, studies, research, design, methodology, critical analysis, interventions, outcomes, data

This commissioned paper is based on a talk given by the author on 26th September 2009 at the national primary care conference of the General Practice Airways Group (GPIAG) held at the University of Warwick, UK. The next national conference of the GPIAG (now the Primary Care Respiratory Society UK) will be held at the Telford International Centre, Shropshire, on 10th – 11th September 2010.

Introduction

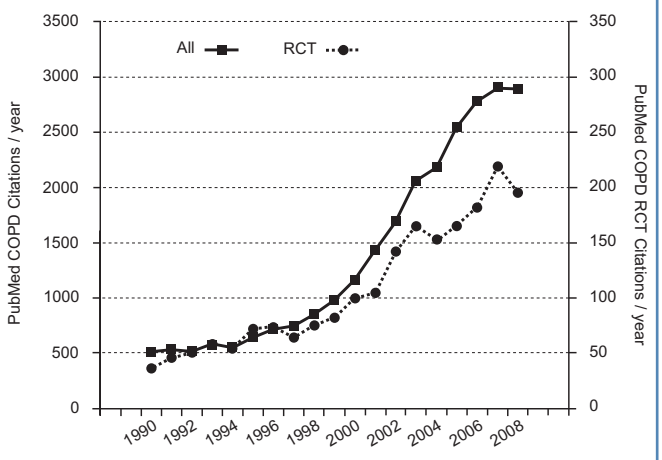
> "*Between measurements based on RCTs and benefit in the community there is a gulf which has been much under-estimated*".
>
> AL Cochrane[1]

The number of publications on the management of COPD has increased enormously in recent years. In 1990 there were just over 500 citations on COPD on PubMed (Medline), and this had risen to nearly 3000 by 2008 (see Figure 1). The number of citations of randomised clinical trials in COPD has risen from just under 50 per year in the early 1990s to over 200 a year in 2008. Translating this evidence base into clinical practice is challenging, and trial designs raise a number of issues about the applicability of the results.

This article will discuss the populations that are included in recent COPD studies, the nature of the interventions, and the importance of 'drop-outs' and 'non-completers'. It will review the relevance of individual outcomes, the incidence of adverse events, the effect of study duration and economic issues around treatments. Finally, it will look at how the results of individual studies can or cannot be put together to develop algorithms and clinical guidelines for the management of COPD.

*** Corresponding author:** Dr David MG Halpin, Consultant Physician & Honorary Senior Clinical Lecturer in Respiratory Medicine, Royal Devon & Exeter Hospital, Exeter, EX2 5DW, UK. Tel: 44 (0)1392 402133 Fax: +44 (0)1392 402828 E-mail: D.M.G.Halpin@ex.ac.uk

Figure 1. Annual PubMed citations for COPD since 1990 (All = all papers on COPD, RCT = randomised controlled trials in COPD)



Is the trial well designed and well conducted?

There is often an assumption that if a trial has been published, then the design, conduct and analysis of the study must be appropriate. Unfortunately this is not always the case. When the quality of studies is assessed using structured tools,[2] many are found to have serious flaws. Studies should have enough participants to be able to reach a negative conclusion (i.e. be adequately powered), including any planned subgroup analyses. The outcome measures should be decided in advance and the analyses of these should be specified before the study is undertaken. This minimises the risk of "data dredging" to find positive results in specific subgroups. The study should be set up in a way that ensures that the researchers are unaware of which patients are receiving which treatment (i.e. allocation concealment and blinding). These aspects of the study design should all be reported in the paper, and failure to do so has been associated with bias in estimating the effectiveness of interventions.[3,4]

All recent studies should conform to the standards of the CONSORT statement[5] and state that they do so. One of the most useful requirements of the CONSORT statement is the recommendation that a diagram of the flow of patients through the study is included. This allows readers to assess what proportion of the potential participants actually took part in the study and the reasons why those who were not enrolled did not participate, as well as the numbers of people who dropped out during the study and the reasons for this. It also allows readers to assess whether the authors have performed an intention-to-treat analysis.[6,7]

Failure to randomise all eligible patients because the study clinicians only offer participation in the trial to patients they considers suitable for the intervention is another important source of potential bias; the inclusion in the CONSORT diagram of the numbers of patients considered for the study as well as the number randomised lets the reader know whether the study has included a representative group of patients or whether they were a highly selected group.

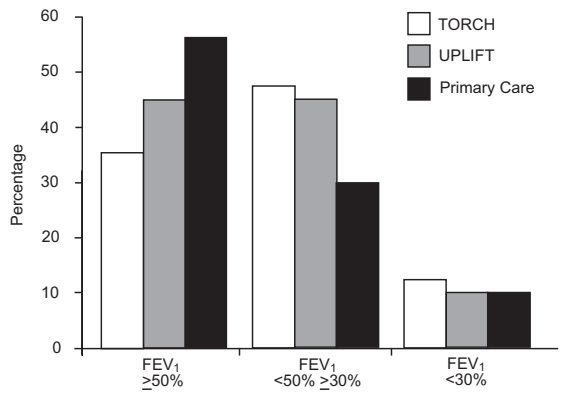
Are trial participants similar to the people you treat?

When reading clinical trial papers or listening to the presentation of the results, it is tempting to skim over the demographics of the people participating in the study. These are often quite similar in different trials: the mean age is often around 65, around 75% of the participants are male, and the mean pre-bronchodilator forced expiratory volume in one second (FEV_1) is often around 45% predicted. Nevertheless, it is important to review this information when considering whether the results of the study are applicable in clinical practice.

All clinical trials have inclusion and exclusion criteria which are designed to recruit a reasonably homogeneous population which will give clear results with little intra-individual variation. Many of the common inclusion criteria specify age ranges which exclude older patients from participating in studies. For example, the TORCH study[8] recruited patients aged 40-80 but excluded patients over the age of 80. This potentially affects the application of the studies in clinical practice. However, even studies which do not include an upper age cut-off (e.g. UPLIFT[9]) include relatively few older people; the mean age of participants in TORCH and UPLIFT, and the age distribution (expressed as standard deviation), were similar – 65.0±8.3 yrs and 64.5±8.3 yrs, respectively. This demonstrates the importance of looking at the demographic details of who was actually studied as well as who it was intended to study.

Similarly, studies often also make stipulations about the severity of the disease, either in terms of the FEV_1 as a percent predicted (pre- or post-bronchodilator) or the frequency of exacerbations over a period prior to entry to the trial. They may also have a requirement for a lack of bronchodilator response. The TORCH study included people with a pre-bronchodilator FEV_1 of less than 60% predicted, and an increase in FEV_1 following 400 mcg of salbutamol of less than 10% of the predicted value for that individual.[8] UPLIFT required a post-bronchodilator (i.e. after 80 mcg ipratropium and 400 mcg salbutamol) FEV_1 of 70% or less of the predicted value, and an FEV_1/FVC ratio of 0.7 or less, but did not have a cut-off for the bronchodilator response.[9] Despite these differences in the lung function criteria, the people recruited were on average similar (mean pre-bronchodilator FEV_1 approximately 44% predicted in TORCH[8] and 39.5% predicted in UPLIFT[9]); however, there are

Figure 2 Distribution of patients by GOLD stage in TORCH, UPLIFT and in primary care in England.



differences in the relative proportion of patients with mild, moderate and severe airflow obstruction defined according to pre-bronchodilator FEV_1 (see Figure 2). This is important, since the UPLIFT study was designed to allow a separate analysis of outcomes in patients with an FEV_1 of 50% predicted or more (as discussed below). Even though UPLIFT had more GOLD stage 2 participants, there were still fewer of these than in a typical primary care population.[10] This may also have an effect on the applicability or generalisability of the study results – particularly in general practice.

Many studies exclude patients with co-morbid illnesses, whereas in practice such people form the bulk of patients that are treated. Studies may also exclude patients considered to be at high risk of dying during the study – especially if it lasts more than 12 months – and most clinical trials exclude patients on long term oxygen therapy. Consideration of the potential implications of excluding these patients is important when interpreting the study results, particularly the adverse event data.

Are the results applicable in patients of all types?

Baseline mediation use and the effect of the trial protocol on this is very important when interpreting the study results. Many patients entering COPD studies are on both short-acting reliever bronchodilator therapy as well as long-acting bronchodilator maintenance therapy and possibly oral or inhaled steroids (see Table 1). Some trial designs (e.g. one of the budesonide/formoterol studies[13] and TORCH[8]) required these medications to be stopped during a run-in period, potentially leading to a deterioration of the patient's clinical condition, whilst others allowed them to be continued at stable doses (e.g. UPLIFT[9]); and some studies included an optimisation phase when additional medication was started in all patients (e.g. the other budesonide/formoterol study[11]). If participants were on maintenance therapy and had this

Table 1. Baseline medication use in COPD trials.

Baseline medication use	TORCH[8]	UPLIFT[9]	Calverley Symbicort[11]	Calverley Roflumilast[12]
LABA	9%	60%	~30%	~49%
LAMA	NR	2%		NR
ICS	19%	62%	48%	~40%
LABA/ICS	28%	NR	NR	NR
Oral steroids	NR	8.4%	NR	NR

stopped at recruitment, the study design can be viewed as a "withdrawal" study; it can be considered that the results of relevance are those in the group who had their therapy stopped rather than those in whom therapy was continued.[14]

Studies which allow patients to continue current mediation during the study should be interpreted as "add-on" studies unless the study is sufficiently large to have statistical power to allow pre-specified analyses of the results in patients receiving different medication at baseline – although such analyses are potentially susceptible to bias. Unless patients are randomised into strata based on medication use, subgroup analysis according to concomitant medication use negates the process of randomisation and it cannot be assumed that the characteristics of patients in the comparison groups are similar. Thus, any conclusions that are reached must be considered hypothesis-generating rather than definitive.

Some studies attempt to assess the effects of therapy in groups of patients with COPD of different severity. This is best done by pre-specifying sub group analyses based on the degree of FEV_1 impairment; for example, UPLIFT specifically recruited enough patients with an FEV_1 of 50% or higher to allow an analysis of the effects in these patients.[15] In some studies, analysis of the results by GOLD stage was not planned prior to the start of the study, but sufficient numbers of patients can sometimes be identified to allow an analysis,[16] and provided the groups are of sufficient size it is unlikely that significant bias will be introduced because of non-random distribution of patients between groups. Alternatively, some studies recruit patients with a narrower FEV_1 range to try and assess benefits in that particular group.[17]

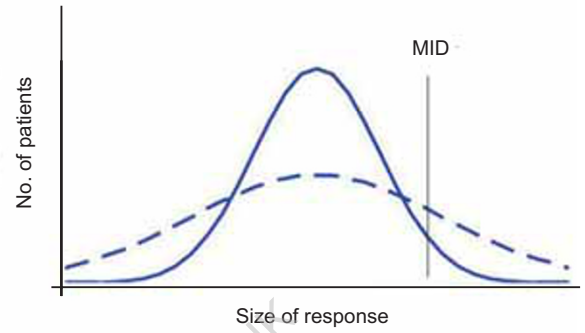
Some studies undertake secondary analyses looking at the effects of the intervention in specific subgroups – e.g. smokers versus ex-smokers[18] – and this may be of considerable relevance to clinicians. However, again there may be biases in the analysis unless it was planned as part of the study design and patients were randomised by the subgroup of interest.

Many studies are not large enough individually to allow sub-group analysis but have a similar trial design, thereby allowing pooled analysis of patient level data or meta-

Table 2. Essential requirements for a meta-analysis. From Egger *et al*.[20]

Meta-analysis should be as carefully planned as any other research project, with a detailed written protocol being prepared in advance
The *a priori* definition of eligibility criteria for studies to be included and a comprehensive search for such studies are central to high quality meta-analysis
The graphical display of results from individual studies on a common scale is an important intermediate step, which allows a visual examination of the degree of heterogeneity between studies
Different statistical methods exist for combining the data, but there is no single "correct" method
A thorough sensitivity analysis is essential to assess the robustness of combined estimates to different assumptions and inclusion criteria

Figure 3 Two hypothetical distributions of treatment effects with the same mean response but differences in the number of people achieving a response greater than the minimum important difference (MID).



analyses by subgroups;[19] but as discussed above such analyses may be affected by bias unless patients were randomised to the sub-groups. There is an increasing trend to use meta-analysis to delve more deeply into data, and Egger *et al*. have identified important points that must be met for a meta-analysis to be sound[20] (see Table 2).

To overcome some of the limitations that result from the design and conduct of randomised controlled trials (RCTs), some studies use an observational design to look at clinical outcomes in "real world" settings. These generally use information from administrative or prescribing databases and have the strength that they include all types of patients (age, sex, smoking status, FEV_1, co-morbidities).[21] Yet interpretation of these studies can be difficult; the outcomes observed may be confounded by the fact that clinicians obviously had a reason to prescribe the therapy and thus there are likely to be systematic differences between patients who received the intervention and those who did not.[22] Although attempts are often made to try to control for these confounders, some of the information upon which the decision was made may not be available in the database and many proponents of evidence-based medicine dismiss observational studies: "If you find that [a] study was not randomised, we'd suggest that you stop reading it and go on to the next article."[23] However, a more recent comparison of the results of observational studies with those of RCTs in 136 reports of 19 different and diverse interventions, found that in most cases the estimates of the treatment effects from observational studies were similar to those found in RCTs.[24] Given the strengths of observational studies, particularly the inclusion of a broader range of patients, their results should not be dismissed out of hand. They may need to be interpreted with caution if their results appear to be at variance with the results of randomised trials, but then it may be that it is the results of the RCT which are misleading for one or more of the reasons discussed above.

How are the data presented?

When reporting outcomes most studies will report a value for an outcome at a particular point in time. This figure usually represents the mean value in the group of patients being studied ('group mean data'), and should be accompanied by an estimate of the range of values observed – for example, 95% confidence intervals, standard error, or standard deviations from the mean value. The range is very important, but in graphical presentation of results, error bars which show the range of values are often omitted, usually for clarity. In doing so, results may appear more important than they really are.

Analysis of group mean data may also sometimes miss important treatment effects. Figure 3 shows the distribution of responses in two hypothetical populations: they could represent the change in St Georges Respiratory Questionnaire (SGRQ) values after six months' treatment with a new bronchodilator. Both groups have the same mean value, and this is below the clinically meaningful difference of 4 units, but more patients in group 2 have a value above this threshold, suggesting that more people have responded. This is balanced by the fact that more people have shown a smaller increase; but it may be misleading to conclude that both treatments are equally effective. The FDA has suggested that this approach may be of particular relevance for patient reported outcomes (PROs):

> *There may be situations where it is more reasonable to characterise the meaningfulness of an individual's response to treatment than a group's response, and there may be interest in characterising an individual patient as a responder to treatment, based upon pre-specified criteria backed by empirically derived evidence supporting the responder definition as a measure of*

benefit. Such examples include categorising a patient as a responder based upon a pre-specified change from baseline on one or more scales; a change in score of a certain size or greater (e.g., a 2-point change on an 8-point scale); or a percent change from baseline."[25]

The key point is that the minimum important difference (MID) that will be used is specified before the trial is undertaken,[26,27] otherwise there is a danger that the researchers will analyse the data retrospectively and use a value chosen to show a difference between groups. The MID for the SGRQ is well established[28] and studies do report differences in the proportion of patients achieving this. For example, a study of the effects of salmeterol/fluticasone versus salmeterol showed that the number of patients with a reduction in SGRQ of 4 points or more was significantly higher in the combination group (41.7% v 30%),[29] and in the UPLIFT study a higher proportion of patients in the tiotropium group than in the placebo group had an improvement of 4 units or more in the SGRQ total scores from baseline at 4 years (45% vs. 36%).[9]

How do non-completers affect interpretation?

Non-completing patients are almost inevitable in trials of COPD interventions that last more than a few weeks. The patients who do not complete the study are usually dropping out because of perceived lack of efficacy, and differential drop-out can itself be seen as a marker of efficacy;[30] however, non-random drop-out introduces another potential source of bias to the analysis of study. Analysis of the baseline characteristics of patients who drop out of placebo groups shows that they generally have more severe disease, whilst those dropping out of the active treatment groups are similar to patients who complete studies. This suggests that active treatment may allow the more physiologically impaired patients to cope better with worsening symptoms and exacerbations and therefore remain in the study.

Non-completers are often patients who deteriorate most quickly during the study.[30] Thus, drop-out from across the treatment groups can lead to a healthier survivor effect and under-estimation of the effect of an intervention, since the outcomes in different treatment limbs converge as sicker patients drop out. Despite the problems they introduce, data from non-completers should be included in the analysis. Standard statistical methods used to analyse these studies (such as negative binomial models) take account of the length of time each individual spends in the trial, and make assumptions such as 'missing data are randomly distributed and patients lost to follow-up are just as likely to have future events as those staying in the study'.[31,32]

If outcomes are recorded for patients who have left the study and the results are analysed on an intention to treat basis, it is important to realise that some patients will start active therapy even though their outcomes will be analysed according to which group they were in originally. For example, in the TORCH study, 44% of people in the placebo group dropped out during the three years of follow up; yet by the end of the follow-up period, over 50% of these patients had started open-label treatment with long-acting β_2-agonists (LABA), inhaled corticosteroids (ICS), or LABA+ICS,[32] but the mortality data for this group was analysed correctly as if they had remained on placebo – potentially leading to underestimation of the magnitude of the benefit of combination therapy over "placebo".

Despite being subject to the potential bias described above, when mortality is an outcome, complete follow-up is very important since patients whose clinical condition is deteriorating are likely to drop out of studies and may die soon after. Unless this information is recorded, a true assessment of the impact on mortality cannot be made. This has been a criticism of the INSPIRE study where complete follow-up was not available,[33] and when mortality rates from this study are compared with those in studies where complete follow-up has been undertaken a considerable difference is seen.[34]

For other outcomes, complete follow-up is generally not possible once a patient has dropped out of a study since these patients do not attend for follow-up visits and further data about their symptoms, exacerbations and lung function is not available. Although it would be desirable to be able to collect this information,[14] the nature of the consent process usually prevents this. Even if trial participants give permission at the beginning of the study for them to be contacted if they drop out of the study, they can withdraw this permission at any time, thus leading to incomplete data.

When outcomes such as exacerbation rates have been monitored in as many participants as possible to the end of the study, even if the subjects have stopped taking the medication to which they were randomised, significant differences are seen between results for the whole study period on an 'intention to treat' basis and results for the time people remain in the study

Table 3. Analyses of the effect of inhaled corticosteroid (ICS) use compared with bronchodilators on the rate of exacerbation over different follow-up periods using the Canadian Optimal Therapy of COPD Trial data. Adapted with permission from Suissa *et al*.[14]

	Exacerbation rate ratio (ICS v Placebo)
Whole period (ITT)	0.83 (95% CI 0.66-1.04)
Prior to withdrawl	0.78 (95% CI 0.61-0.99)
After drug discontinuation	1.23 (95% CI 0.78-1.95)

and for the period after they discontinued randomised drug therapy (see Table 3).

For all these reasons it is important to consider what proportion of patients dropped out during the treatment period and how data relating to these patients were analysed when considering the relevance of the study results to routine practice.

Are you using the same intervention?

RCTs use specific drugs at fixed doses in certain devices. The results of the study, therefore, can really only be translated into practice if the same drug dose and device are used. In practice, many clinicians assume that drugs within a certain class will have similar efficacy, and often dosages are reduced on cost grounds or because of concerns about side effects. There is little intellectual justification for this. Whilst it may be possible to assume that the effects of a treatment in one population may be applicable in a similar but slightly different population, there is no reason to assume that lower doses will produce the same effect unless studies have shown this. For salmeterol/fluticasone, the 3-year TORCH study and 1-year Tristan study used a dose of 500/50 twice daily. A 24-week study showed that 250/50 twice daily also significantly improved pre- and post-dose FEV_1 compared to salmeterol alone or placebo,[35] but the mean FEV_1 of patients enrolled in this study was lower than that in the higher dose longer studies (33% predicted v 44% in TORCH & Tristan) making it difficult to make direct comparisons about the dose response.

Two more recent 52-week studies have shown that the 250/50 dose twice daily reduced exacerbations compared to salmeterol – but these studies included an optimisation phase and did not include a placebo arm, so again it is difficult to make firm conclusions about the comparative effects of a lower dose.[36,37]

Patients and clinicians would like reassurance that the dose of medication they are using is optimal in terms of risks and benefits, but unless dose-ranging studies have been performed, the doses used in the trials delivered using the same devices are the only ones that can be recommended according to evidence-based medicine.

Similarly, unless the pharmacological properties of different drugs in the same class are very similar (e.g. mode of action, speed of onset, duration of action, etc.) it is also inappropriate to assume that they will produce similar clinical benefits.

Are you interested in the same outcomes?

A range of outcome measures have been used in studies of therapy in COPD. These include physiological measures such as FEV_1 and inspiratory capacity, and symptom measures such as breathlessness and cough scores. Quality of life is usually measured using a questionnaire, and exacerbation data may be recorded on the basis of symptom diaries or retrospectively at study visits. Recent long-term studies have also collected data on mortality. Some outcomes such as fatigue are important to patients but are difficult to measure, and trials may miss important benefits because of this.

When considering the effects of an intervention it is important to consider whether the outcomes that are desired in practice are the ones that were used in the clinical trial. In many cases they are, but sometimes it is assumed that improvements in clinically relevant outcomes can be inferred from changes in surrogate end points, whereas this may not be the case.

Are adverse advents important?

Drug interventions can be associated with predictable side effects such as dry mouth or tremor, but sometimes unexpected side effects appear in clinical trials. This is the case with the adverse event reporting of non-fatal pneumonia in the TORCH study.[38] Similar effects have been found in other studies that used fluticasone.[29,36,37,39] They did not appear to be seen in studies that use budesonide, and this has been confirmed in a more detailed patient-level pooled analysis.[40] Despite the apparent increase in the risk of pneumonia there was no associated increase in death from pneumonia in these studies.

On the other hand TORCH was very reassuring about the effect of ICS on the risk of osteoporosis. Although osteoporosis is common in people with COPD, three years' treatment with fluticasone did not increase the risk of fractures or reduce bone mineral density compared to placebo.[41] Similarly there was no increased risk of developing cateracts in people treated with fluticasone.[8]

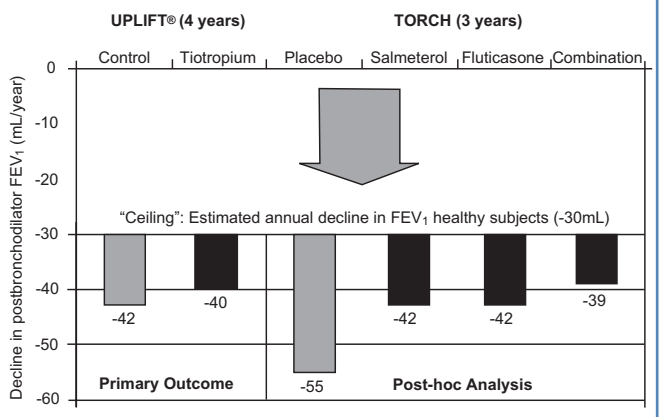
On the basis of a meta-analysis there had also been concerns that long term use of salmeterol alone was associated with an increased risk of adverse events including death.[42] However, TORCH has clearly shown that salmeterol monotherapy was not associated with increased mortality or other adverse events.

Assessing the importance of adverse advents is often difficult, and the benefits of therapy need to be weighed against potential adverse events. One way to do this is comparing the numbers needed to treat (NNT) to achieve a good outcome, and the numbers needed to harm (NNH) for adverse effects.[43]

Can the impossible be achieved?

When looking at the outcomes of studies it is important to realise that in many cases the extent to which outcomes can be improved is limited. For example, the rate of decline of lung function cannot be slowed any more than the normal rate of decline. There may also be ceiling effects for some outcomes, such as slowing the rate of decline in FEV_1, so that adding further drugs has little or no benefit. Thus, in UPLIFT it is possible that tiotropium was unable to reduce the decline in

Figure 4 Change in rate of decline in FEV_1 in UPLIFT and TORCH, including ceiling effect (rate of decline in healthy individuals). Data from the individual trials have been placed on the same axes for illustrative purposes only and do not represent directly comparable data between the trials. Reproduced with permission from Miravitlles & Anzuetto 2009.[44]



FEV_1 further because the majority of the patients in the comparison group were on LABA/ICS therapy and continued this during the trial (see Figure 4).[44]

What is clinically relevant?

Studies show that current therapy can significantly improve symptoms, reduce exacerbation rates, improve quality of life and slow the decline in FEV_1, as well as reducing mortality. These effects are all statistically significant, but knowing what is clinically important is more difficult. As mentioned above, the minimal important difference (MID) has been defined as "the smallest difference in score in the outcome of interest that informed patients or informed proxies perceive as important, either beneficial or harmful, and which would lead the patient or clinician to consider a change in management."[45] A number of MIDs for outcomes in COPD have been proposed, and the evidence for these has been summarised in the report of the ATS/ERS task force on outcome measures in COPD[33] (see Table 4).

MIDs can be estimated in different ways. They may be developed by consensus of experts and patients, they may be derived from statistical analysis of the distribution of the outcome measure (e.g. the MID has sometimes been defined as half the standard deviation (SD) of the measurement, or the standard error (SE) of the measurement), they may be defined from the distribution of changes in the outcome measure produced by an intervention (e.g. the MID has been defined on the basis of the effect size; the average change divided by the baseline SD), or the MID may defined using external- or anchor-based methods, which compare changes in the outcome of interest to other clinically important outcomes.[52,53]

It is important to realise that MIDs are dependent on the clinical characteristics of the patients in which they were determined, and a different MID may apply in different patient groups – e.g. more severe patients or people of different age.[54]

Does the effect persist over a clinically meaningful period?

Depending on the outcome that is being assessed, some intervention studies will last only hours whereas others will last up to four years. Short-term studies have the advantage of very good data collection; however, it can be difficult to be confident that the results can be translated into routine clinical practice when managing chronic disease. Studies lasting only hours, days or weeks are important to establish

Table 4. Suggested minimal important differences (MIDs) of some commonly used outcomes in COPD trials (modified from Cazzola *et al.*[33]).

Outcome measure	Suggested MID	Reference
Respiratory-specific health status and HRQoL		
St George's Respiratory Questionnaire	4 units	(46)
Chronic Respiratory Questionnaire	0.5 units for the average score on each domain	(47)
Dyspnoea		
Transition dyspnoea index	1 unit	(48)
Lung function		
FEV_1	100–140 mL	(49)
Exercise		
6-min walking test	37–71 m	(50)
Other		
Exacerbations of COPD	1 exacerbation per yr, 22% change	(51)

pharmacological properties of a drug but to look at clinically meaning outcomes in COPD it is important that studies last at least three months and ideally six months. When looking at longer term changes in the disease – such as the rate of decline in FEV_1 or mortality studies – several years are required.

There is a trade-off between the accuracy with which outcomes can be monitored over long periods of time and the problems caused by drop-outs as patients deteriorate. Thus, although it may seem attractive to undertake very long-term studies of outcomes in COPD over periods such as 10 years, in practice the number of drop-outs that would occur during such a study are likely to render the results uninterpretable.

Economic issues

The cost of an intervention is often not taken into account in the primary study. There may be secondary health economic analyses published as separate manuscripts which undertake detailed analysis of costs, either on a 'bottom-up' or 'top-down' basis. The methodological quality of these studies is very often high; however, sometimes the significance of the primary study data upon which the health economic analyses are based is questionable, and this fact may get lost in the secondary manuscript.[55]

Cost effectiveness analyses are generally the most relevant economic analyses.[56] They compare the costs of an intervention with the outcomes it produces measured in universal units. Effectiveness is often measured in units such as 'life years gained' and these can be adjusted for the quality of life experienced during the gained years to give a 'quality-adjusted life year' (QALY). Combinations of drugs may be more effective than therapy with single agents – but cost effectiveness analyses may show that the gains are achieved at a cost which is prohibitive.

A concurrent health economic analysis of the OPTIMAL study found that although it was more effective, the cost per QALY gained with triple therapy with tiotropium, salmeterol and fluticasone compared to tiotropium alone was CAN$243,180.[57] This is well above the threshold used by the UK National Institute for Clinical Effectiveness (NICE) to assess cost effectiveness (£20-30,000 per QALY[58]). A secondary economic analysis of the TORCH trial[59] estimated that salmeterol/fluticasone had an overall cost per QALY of $43,600 compared to placebo, SFC had a lower cost effectiveness ratio than either salmeterol or fluticasone alone ($197,000 and $78,000 per QALY respectively), and in this case the economic analysis supports the clinical evidence that combination therapy is the better treatment.

The cost of an intervention and the costs it saves may well depend on the country in which the study was performed or modeled – since healthcare systems significantly affect the cost of an intervention and the potential savings from events such as hospitalisation. Although a model may suggest that a drug is cost effective in a certain health economy it cannot automatically be assumed that it will be cost effective in all health economies.[60] In the secondary economic analysis of the TORCH trial,[59] there were geographical variations in the cost effectiveness of salmeterol/fluticasone, with an estimate of cost per QALY of $24,200 in Western Europe and $77,100 in the USA.

It is important to remember, however, that even if a treatment is cost effective, there are challenges for many payors (including the NHS in the UK) in deciding whether it is affordable or not within a fixed healthcare budget.[61]

What do trials tells us about the sequencing of therapy?

The best RCTs answer only very specific questions and cannot usually help determine the sequencing of therapies for an individual patient. A number of different RCTs may all show the effectiveness of particular interventions, but the benefits of combining interventions or introducing one intervention before another are not usually tested in studies.

RCTs do not usually look at the outcomes achieved by different models of care rather than individual specific interventions. There are also few if any trials that compare pharmacological interventions with non-pharmacological interventions such as pulmonary rehabilitation. This leaves guideline developers struggling to put together algorithms for COPD management based on the individual trials, and in this situation it is often consensus and clinical experience that determines the sequence rather than evidence.

Conclusion

It is important to realise that most clinical trials only look at pharmacological interventions and do not include non-pharmacological interventions such as smoking cessation or pulmonary rehabilitation. The more recent longer clinical trials almost certainly give information that is more relevant to clinical practice than earlier shorter trials. The newer studies also include a wider range of patients who are more representative of those seen in clinical practice. On the basis of the results of these studies it is clear that long-acting drugs offer greater benefits than short-acting drugs and are generally more cost effective, largely because of reductions in hospitalisation rates. Multiple combinations offer additional benefits over individual treatments but the cost effectiveness of these therapies is less clear. Although adverse events are recorded in COPD trials, in general the benefits of therapy on quality of life and reduced exacerbation rates outweigh these risks.

Conflict of interest declaration

Dr Halpin has received sponsorship to attend international meetings, and honoraria for lecturing, attending advisory boards and preparing educational materials from, AstraZeneca, Boehringer Ingelheim, Chiesis, Nycomed, GlaxoSmithKline, and Pfizer. His department has received research funding from AstraZeneca.

References

1. Cochrane AL. Effectiveness and Efficiency: random reflections on Health Services. London: Nuffield Provincial Hospitals Trust; 1972.
2. Jüni P, Altman DG, Egger M. Assessing the quality of randomised controlled trials. In: Egger M, Davey Smith G, Altman DG, editors. Systematic reviews in health care: meta-analysis in context. London: BMJ Books; 2001. p. 87-121. http://dx.doi.org/10.1002/9780470693926.ch5
3. Schulz K, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;**273**:408-12.
4. Moher D, Pham B, Jones A, *et al*. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;**352**:609-13. http://dx.doi.org/10.1016/S0140-6736(98)01085-X
5. Altman DG, Schulz KF, Moher D, *et al.* The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;**134**:663-94.
6. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 1999;**319**:670-4.
7. Lee YJ, Ellenberg J, H, Hirtz DG, Nelson KB. Analysis of clinical trials by treatment actually received: is it really an option? *Stat Med* 1991;**10**:1595-605. http://dx.doi.org/10.1002/sim.4780101011
8. Calverley PM, Anderson JA, Celli B, *et al*. Salmeterol and fluticasone propionate and survival in chronic obstructive pulmonary disease. *N Engl J Med* 2007;**356**(8):775-89. http://dx.doi.org/10.1056/NEJMoa063070
9. Tashkin DP, Celli B, Senn S, *et al.* A 4-year trial of tiotropium in chronic obstructive pulmonary disease. *N Engl J Med* 2008;**359**(15):1543-54. http://dx.doi.org/10.1056/NEJMoa0805800
10. Strong M, South G, Carlisle R. The UK Quality and Outcomes Framework pay-for-performance scheme and spirometry: rewarding quality or just quantity? A cross-sectional study in Rotherham, UK. *BMC Health Serv Res* 2009;**9**:108. http://dx.doi.org/10.1186/1472-6963-9-108
11. Calverley PM, Boonsawat W, Cseke Z, Zhong N, Peterson S, Olsson H. Maintenance therapy with budesonide and formoterol in chronic obstructive pulmonary disease. *Eur Respir J* 2003;**22**(6):912-19. http://dx.doi.org/10.1183/09031936.03.00027003
12. Calverley PM, Rabe KF, Goehring UM, Kristiansen S, Fabbri LM, Martinez FJ. Roflumilast in symptomatic chronic obstructive pulmonary disease: two randomised clinical trials. *Lancet* 2009;**374**(9691):685-94. http://dx.doi.org/10.1016/S0140-6736(09)61255-1
13. Szafranski W, Cukier A, Ramirez A, *et al.* Efficacy and safety of budesonide/formoterol in the management of chronic obstructive pulmonary disease. *Eur Respir J* 2003;**21**(1):74-81. http://dx.doi.org/10.1183/09031936.03.00031402
14. Suissa S, Ernst P, Vandemheen KL, Aaron SD. Methodological issues in therapeutic trials of COPD. *Eur Respir J* 2008;**31**(5):927-33. http://dx.doi.org/10.1183/09031936.00098307
15. Decramer M, Celli B, Kesten S, Lystig T, Mehra S, Tashkin DP. Effect of tiotropium on outcomes in patients with moderate chronic obstructive pulmonary disease (UPLIFT): a prespecified subgroup analysis of a randomised controlled trial. *Lancet* 2009;**374**(9696):1171-8.
16. Jenkins CR, Jones PW, Calverley PM, *et al.* Efficacy of salmeterol/fluticasone propionate by GOLD stage of chronic obstructive pulmonary disease: analysis from the randomised, placebo-controlled TORCH study. *Respir Res* 2009;**10**:59.
17. Johansson G, Lindberg A, Romberg K, Nordstrom L, Gerken F, Roquet A. Bronchodilator efficacy of tiotropium in patients with mild to moderate COPD. *Prim Care Resp J* 2008;**17**(3):169-75. http://dx.doi.org/10.3132/pcrj.2008.00037
18. Tashkin DP, Celli B, Kesten S, Lystig T, Mehra S, Decramer M. Long-term efficacy of tiotropium in relation to smoking status in the UPLIFT trial. *Eur Respir J* 2010;**35**(2):287-94.
19. Halpin D, Menjoge S, Viel K. Patient-level pooled analysis of the effect of tiotropium on COPD exacerbations and related hospitalisations. *Prim Care Resp J* 2009;**18**(2):106-13. http://dx.doi.org/10.4104/pcrj.2009.00017
20. Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. *BMJ* 1997;**315**(7121):1533-7.
21. Griffin J, Lee S, Caiado M, Kesten S, Price D. Comparison of tiotropium bromide and combined ipratropium/salbutamol for the treatment of COPD: a UK General Practice Research Database 12-month follow-up study. *Prim Care Resp J* 2008;**17**(2):104-10. http://dx.doi.org/10.3132/pcrj.2008.00024
22. Chalmers TC, Celano P, Sacks HS, Smith H. Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983;**309**:1358-61.
23. Sackett DL, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine. How to practice and teach EBM. London: Churchill-Livingstone; 1997.
24. Benson K, Hartz AJ. A Comparison Of Observational Studies And Randomized, Controlled Trials. *N Engl J Med* 2000;**342**:1878-86. http://dx.doi.org/10.1056/NEJM200006223422506
25. FDA. Draft Guidance or Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2006. Washington: U.S. Department of Health and Human Services, Food and Drug Administration; 2006.
26. Senn S. Disappointing Dichotomies. *Pharmaceutical Statistics* 2003;**2**:239-40. http://dx.doi.org/10.1002/pst.90
27. Snapinn SM, Jiang Q. Responder analyses and the assessment of a clinically relevant treatment effect. *Trials* 2007;**8**:31. http://dx.doi.org/10.1186/1745-6215-8-31
28. Jones PW. Interpreting thresholds for a clinically significant change in health status in asthma and COPD. *Eur Respir J* 2002;**19**(3):398-404. http://dx.doi.org/10.1183/09031936.02.00063702
29. Kardos P, Wencker M, Glaab T, Vogelmeier C. Impact of salmeterol/fluticasone propionate versus salmeterol on exacerbations in severe chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2007;**175**(2):144-9. http://dx.doi.org/10.1164/rccm.200602-244OC
30. Calverley PM, Spencer S, Willits L, Burge PS, Jones PW. Withdrawal from treatment as an outcome in the ISOLDE study of COPD. *Chest* 2003;**124**(4):1350-6. http://dx.doi.org/10.1378/chest.124.4.1350
31. Keene ON, Jones MR, Lane PW, Anderson J. Analysis of exacerbation rates in asthma and chronic obstructive pulmonary disease: example from the TRISTAN study. *Pharm Stat* 2007;**6**(2):89-97
32. Keene ON, Vestbo J, Anderson JA, *et al*. Methods for therapeutic trials in COPD: lessons from the TORCH trial. *Eur Respir J* 2009;**34**(5):1018-23. http://dx.doi.org/10.1183/09031936.00122608
33. Cazzola M, MacNee W, Martinez FJ, *et al*. Outcomes for COPD pharmacological trials: from lung function to biomarkers. *Eur Respir J* 2008;**31**(2):416-69. http://dx.doi.org/10.1183/09031936.00099306
34. Suissa S. Methodologic shortcomings of the INSPIRE randomized trial. *Am J Respir Crit Care Med* 2008;**178**(10):1090-1; author reply 1-2.
35. Hanania NA, Darken P, Horstman D, *et al*. The efficacy and safety of fluticasone propionate (250 microg)/salmeterol (50 microg) combined in the Diskus inhaler for the treatment of COPD. *Chest* 2003;**124**(3):834-43. http://dx.doi.org/10.1378/chest.124.3.834
36. Ferguson GT, Anzueto A, Fei R, Emmett A, Knobil K, Kalberg C. Effect of fluticasone propionate/salmeterol (250/50 microg) or salmeterol (50 microg) on COPD exacerbations. *Respir Med* 2008;**102**(8):1099-108. http://dx.doi.org/10.1016/j.rmed.2008.04.019
37. Anzueto A, Ferguson GT, Feldman G, *et al.* Effect of fluticasone propionate/salmeterol (250/50) on COPD exacerbations and impact on patient outcomes. *COPD* 2009;**6**(5):320-9. http://dx.doi.org/10.1080/15412550903140881
38. Crim C, Calverley PM, Anderson JA, *et al*. Pneumonia risk in COPD patients

receiving inhaled corticosteroids alone or in combination: TORCH study results. *Eur Respir J* 2009;**34**(3):641-7. http://dx.doi.org/10.1183/09031936.00193908

39. Wedzicha JA, Calverley PM, Seemungal TA, Hagan G, Ansari Z, Stockley RA. The prevention of chronic obstructive pulmonary disease exacerbations by salmeterol/fluticasone propionate or tiotropium bromide. *Am J Respir Crit Care Med* 2008;**177**(1):19-26. http://dx.doi.org/10.1164/rccm.200707-973OC
40. Sin DD, Tashkin D, Zhang X, *et al*. Budesonide and the risk of pneumonia: a meta-analysis of individual patient data. *Lancet* 2009;**374**(9691):712-19. http://dx.doi.org/10.1016/S0140-6736(09)61250-2
41. Ferguson GT, Calverley PM, Anderson JA, *et al*. Prevalence and progression of osteoporosis in patients with COPD: results from the towards a revolution in COPD health study. *Chest* 2009;**136**(6):1456-65. http://dx.doi.org/10.1378/chest.08-3016
42. Salpeter SR, Buckley NS, Salpeter EE. Meta-analysis: anticholinergics, but not beta-agonists, reduce severe exacerbations and respiratory mortality in COPD. *J Gen Intern Med* 2006;**21**(10):1011-19. http://dx.doi.org/10.1111/j.1525-1497.2006.00507.x
43. Halpin DM. Evaluating the effectiveness of combination therapy to prevent COPD exacerbations: the value of NNT analysis. *Int J Clin Pract* 2005;**59**(10):1187-94. http://dx.doi.org/10.1111/j.1368-5031.2005.00664.x
44. Miravitlles M, Anzueto A. Insights into interventions in managing COPD patients: lessons from the TORCH and UPLIFT studies. *Int J Chron Obstruct Pulmon Dis* 2009;**4**(1):185-201.
45. Schunemann HJ, Guyatt GH. Commentary--goodbye M(C)ID! Hello MID, where do you come from? *Health Serv Res* 2005;**40**(2):593-7. http://dx.doi.org/10.1111/j.1475-6773.2005.0k375.x
46. Jones PW. St. George's Respiratory Questionnaire: MCID. *COPD* 2005;**2**(1):75-9. http://dx.doi.org/10.1081/COPD-200050513
47. Schunemann HJ, Puhan M, Goldstein R, Jaeschke R, Guyatt GH. Measurement properties and interpretability of the Chronic respiratory disease questionnaire (CRQ). *COPD* 2005;**2**(1):81-9. http://dx.doi.org/10.1081/COPD-200050651
48. Mahler DA, Witek TJ, Jr. The MCID of the transition dyspnea index is a total score of one unit. *COPD* 2005;**2**(1):99-103. http://dx.doi.org/10.1081/COPD-200050666
49. Donohue JF. Minimal clinically important differences in COPD lung function. *COPD* 2005;**2**(1):111-24. http://dx.doi.org/10.1081/COPD-200053377
50. Wise RA, Brown CD. Minimal clinically important differences in the six-minute walk test and the incremental shuttle walking test. *COPD* 2005;**2**(1):125-9. http://dx.doi.org/10.1081/COPD-200050527
51. Calverley PM. Minimal clinically important difference--exacerbations of COPD. *COPD* 2005;**2**(1):143-8. http://dx.doi.org/10.1081/COPD-200050647
52. Lydick E, Epstein RS. Interpretation of quality of life changes. *Qual Life Res* 1993;**2**(3):221-6. http://dx.doi.org/10.1007/BF00435226
53. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 2003;**56**(5):395-407. http://dx.doi.org/10.1016/S0895-4356(03)00044-1
54. Santanello NC, Zhang J, Seidenberg B, Reiss TF, Barber BL. What are minimal important changes for asthma measures in a clinical trial? *Eur Respir J* 1999;**14**(1):23-7.
55. Halpin DM. Health economics of chronic obstructive pulmonary disease. *Proc Am Thorac Soc* 2006;**3**(3):227-33. http://dx.doi.org/10.1513/pats.200507-072SF
56. Robinson R. Cost-effectiveness analysis. *BMJ* 1993;**307**(6907):793-5. http://dx.doi.org/10.1136/bmj.307.6907.793
57. Najafzadeh M, Marra CA, Sadatsafavi M, *et al*. Cost effectiveness of therapy with combinations of long acting bronchodilators and inhaled steroids for treatment of COPD. *Thorax* 2008;**63**(11):962-7. http://dx.doi.org/10.1136/thx.2007.089557
58. McCabe C, Claxton K, Culyer AJ. The NICE cost-effectiveness threshold: what it is and what that means. *Pharmacoeconomics* 2008;**26**(9):733-44. http://dx.doi.org/10.2165/00019053-200826090-00004
59. Briggs A, Glick H, Lozano-Ortega G, *et al*. Is treatment with ICS and LABA cost-effective for COPD? Multinational economic analysis of the TORCH study. *Eur Respir J* 2009;**35**(3):532-9.
60. O'Brien BJ. A tale of two (or more) cities: geographic transferability of pharmacoeconomic data. *Am J Manag Care* 1997;**3** Suppl:S33-9.
61. Cleary SM, McIntyre D. Affordability--the forgotten criterion in health-care priority setting. *Health Econ* 2009;**18**(4):373-5. http://dx.doi.org/10.1002/hec.1450