Duration of antibiotic treatment and symptom recovery in community-acquired pneumonia
El Moussaoui, R.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Large variability in outcome measures in community-acquired pneumonia (CAP) antibiotic treatment trials – a systematic review

Rachida el Moussaoui¹, Corianne A. de Borgie², Brent C. Opmeer², Patrick M. Bossuyt², Peter Speelman¹, Jan M. Prins¹

¹ Department of Internal Medicine, Division of Infectious Diseases, Tropical Medicine, and AIDS and ² Department of Clinical Epidemiology and Biostatistics, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands.

Submitted
Abstract

Background - We have carried out a systematic review to analyze what criteria are used in randomized controlled trials (RCT’s) to evaluate antibiotics for the treatment of community-acquired pneumonia (CAP), and whether these criteria satisfy the quality standards that should be expected of RCT’s.

Methods - Randomized controlled CAP trials published in English language between 1993 and 2003 were identified. We classified the criteria used to assess clinical response according to the following parameters: signs and symptoms, chest X-ray abnormalities, death by the primary diagnosis, discontinuation of therapy because of adverse events, indeterminate, recurrence/"relapse", new antibiotics needed and temperature.

Results - 28 studies meeting the criteria were identified. The above mentioned outcome parameters were included in respectively 93%, 57%, 32%, 7%, 64%, 50%, 39% and 14% of the studies. Of these studies, only four described what they meant by 'signs and symptoms'. The time point for the post treatment evaluation varied from immediately after treatment to 21 days after treatment.

Conclusions - We conclude that the parameters used to measure outcome were generally based on subjective interpretation, and are therefore prone to interobserver variability, and also lacked standardization, which is important for cross-study comparisons. We suggest that validated patient-based outcome measures, such as the CAP-score and the CAP-sym, should be used to establish clinical effectiveness in RCT’s. All cause 30-day mortality or death by the primary diagnosis, and discontinuation of study treatment because of side-effects or worsening after initial improvement should also be primary endpoints, as these parameters are relatively less susceptible to subjective interpretation.
Introduction

Although advances in the management of CAP have lowered mortality and morbidity rates, innovative therapies are still necessary due to the rapid rise in antimicrobial drug resistance among common respiratory pathogens and the less than optimal tolerability of some existing drugs. Before these newly developed drugs are introduced into clinical practice their effectiveness should be established in clinical trials. Accurate, standardized and validated outcome measures are necessary to allow a systematic comparison of the clinical outcome between the treatment strategies and across controlled clinical trials.

The only available guideline for measuring clinical response in CAP is the one proposed several years ago by the Infectious Diseases Society of America (IDSA): The “Chow criteria” for the evaluation of new anti-infective drugs for the treatment of respiratory tract infections. According to these criteria clinical cure is defined as complete resolution of all signs and symptoms of pneumonia and improvement or lack of progression of all abnormalities on the chest radiograph (table 1). Recent practice guidelines reaffirmed the importance of clinical signs and symptoms plus radiographic findings in the initial diagnosis as well as the evaluation of subsequent response to therapy in CAP.

Several recently published studies, however, have demonstrated that a substantial number of patients with CAP still experience pneumonia-related symptoms 30 days after cessation of antibiotic treatment. These findings indicate that clinical cure as defined by the Chow criteria cannot be established at the end of treatment, since at this moment most patients will report one or more of these symptoms. Another shortcoming in these criteria is that ‘signs and symptoms’ are not well defined and have certainly not been validated. As a consequence, the Chow criteria are diffuse and can be interpreted in many different ways by the clinicians, and the same holds true for the criteria offered in the practice guidelines.

As many of studies on CAP are initiated and/or sponsored by the pharmaceutical industry, we think it is necessary to agree upon the design of such studies, in order to be able to judge them properly. We have carried out a literature review to find out which criteria have been used in clinical CAP trials during the past ten years to assess treatment effectiveness. We have limited our search to randomized controlled trials (RCT’s). We also investigated at what time point in the trials the assessment of clinical response was carried out.

<table>
<thead>
<tr>
<th>Clinical response</th>
<th>Chow criteria of clinical response</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cure</strong></td>
<td>Complete resolution of all signs and symptoms of pneumonia and improvement or lack of progression of all abnormalities on the chest radiograph</td>
</tr>
<tr>
<td><strong>Failure</strong></td>
<td>Any of the following conditions: Persistence or progression of all signs and symptoms of pneumonia after 3-5 days of therapy. Development of new pulmonary or extrapulmonary clinical findings consistent with active infection. Persistence or progression of radiographic abnormalities Death due to pneumonia An inability to complete the study because of adverse events.</td>
</tr>
<tr>
<td><strong>Indeterminate</strong></td>
<td>Indicates that extenuating circumstances preclude classification as cure or failure.</td>
</tr>
</tbody>
</table>

Table 1
Chapter 8

Methods

Literature search
We searched the databases of the National Library of Medicine through MEDLINE for randomized, controlled CAP trials in humans, published in the English language literature between 1993 and 2003, using the terms “community-acquired infections” and “pneumonia” both as medical subject headings and as text words. We combined both terms. We searched for the medical subject heading “antibiotics” and the text words “antibiotics” and “antibiotic”. We combined those three terms. In addition, we searched for the subject medical heading “double blind method” and for the term “random allocation” as a subject medical heading or as text word and for the term “placebos” also as a subject medical heading or as text word and for the publication type “controlled clinical trial”. We combined those four terms. By combining the three search strategies we were left with 90 potentially eligible studies. We then reviewed the abstracts in a standardized way. Studies including other community-acquired infections, nosocomial or childhood pneumonia were not included. Studies meeting the criteria, but with a total patient population smaller than 50 patients were excluded. A manual search of the bibliographies of relevant papers was done to identify additional eligible publications. In total, 28 studies met these inclusion criteria.

Data extraction
Two reviewers independently reviewed the 28 studies left. To summarize and describe the included studies we extracted the following study characteristics: the year of publication, the study location, the number of included patients, the type and dosage of antibiotics used, the duration of treatment, the moment of evaluation, cure percentages, recurrence rate, and the mode of efficacy analysis: intention to treat (ITT) or per protocol (PP) or both.

Second, we extracted the criteria used to measure clinical response. To compare the various trials, we classified these criteria used to assess clinical response according to the following parameters: 1. signs and symptoms, 2. radiographic chest abnormalities, 3. death by the primary diagnosis, 4. discontinuation of therapy because of adverse events, 5. indeterminate response, 6. recurrence/ “relapse”, 7. new antibiotics needed and 8. temperature. The first five parameters were suggested by Chow. The remaining three parameters have been used as part of clinical response assessment in several trials.

‘Signs and symptoms’ are important features in the assessment of clinical outcome. We, therefore, examined what signs and symptoms were used and how they were rated.

Results

Study characteristics
Characteristics of the 28 studies included are shown in table 2. The number of included patients varied between 86 and 1137 patients, while the duration of treatment ranged from 3 to 14 days. Except for one study9, the percentage of cure was reported in all studies as ITT analysis, PP analysis, or both. The data in table 2 show that the cure rate for the same drug can vary across studies. For example, the reported cure rate of sparfloxacin (four studies, equal dosages) varies from 89% to 95% in the PP analysis for the first assessment period, while the cure rate during the late assessment period in the ITT analysis shows an even greater variability: from 69% to 82%.
### Table 2 Characteristics of the 28 studies included in the analysis of the definition of clinical response

<table>
<thead>
<tr>
<th>Author</th>
<th>Year of publication</th>
<th>Sample size</th>
<th>Treatment</th>
<th>Duration of assessment of TOC* in days</th>
<th>Cure % of early follow-up</th>
<th>Moment of evaluation of Late follow-up in days</th>
<th>Cure % of late follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>2002</td>
<td>502</td>
<td>eritapenem, ceftazidime</td>
<td>≥ 3 7-14 after EOT†</td>
<td>92 (EP) and 85 (ITT)</td>
<td>21-28 after EOT Only relapse was assessed</td>
<td>21-28 after EOT 84 (PP) 74 (PP)</td>
</tr>
<tr>
<td>22</td>
<td>2002</td>
<td>628</td>
<td>moxifloxacin, amoxicillin-clavulananate, with or without clarithromycin</td>
<td>7-14 5-7 after EOT</td>
<td>93 (PP) 85 (PP)</td>
<td>21-28 after EOT 84 (PP) 74 (PP)</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>2002</td>
<td>344</td>
<td>amoxicillin-clavulananate, 2x2000</td>
<td>10 Day 28-35</td>
<td>92 (PP) and 81 (ITT)</td>
<td>31-36 92 (EP) and 81 (ITT) 85 (EP) and 73 (ITT)</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>2002</td>
<td>404</td>
<td>telithromycine, amoxicillin</td>
<td>10 Day 17-24</td>
<td>95 (EP) and 86 (ITT) 90 (EP) and 79 (ITT)</td>
<td>21-28 after EOT 94 (EP) and 87 (ITT)</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>2002</td>
<td>364</td>
<td>eritapenem, ceftazidime</td>
<td>≥ 3 7-14 after EOT</td>
<td>92 (EP) and 86 (ITT) 94 (EP) and 87 (ITT)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>2002</td>
<td>176</td>
<td>clarithromycin, trovafloxacin</td>
<td>7 14-21 after EOT</td>
<td>87 (EP) 95 (EP)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>2001</td>
<td>411</td>
<td>moxifloxacin, amoxicillin</td>
<td>10 3-5 after EOT</td>
<td>92 (PP) and 87 (ITT) 90 (PP) and 82 (ITT)</td>
<td>21-28 after EOT 89 (PP) and 77 (ITT) 89 (PP) and 78 (ITT)</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>2001</td>
<td>678</td>
<td>moxifloxacin, clarithromycin</td>
<td>10 3-5 after EOT</td>
<td>94 (PP) 94 (PP)</td>
<td>21-28 after EOT 91 (PP) 93 (PP) 92 (PP)</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>2000</td>
<td>169</td>
<td>azithromycin, ceftriaxone, or erythromycin</td>
<td>7-10 EOT</td>
<td>91 (EP) and 75 (ITT) 91 (EP) and 83 (ITT)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>2000</td>
<td>283</td>
<td>Gatifloxacin ≤→ gatifloxacin or ceftazidime, or erythromycin, or clarithromycin</td>
<td>? Day 7-14</td>
<td>97 (EP) 91 (EP)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>2000</td>
<td>86</td>
<td>clarithromycin, roxithromycin, erythromycin</td>
<td>10 2 after EOT</td>
<td>89 82 73</td>
<td>15-45 after EOT Unknown</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>1999</td>
<td>342</td>
<td>sparfloxacin, clarithromycin</td>
<td>10 10 days after EOT</td>
<td>89 (PP) and 80 (ITT) 89 (PP) and 83 (ITT)</td>
<td>28 days after EOT 89 (PP) and 76 (ITT) 84 (PP) and 77 (ITT)</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>1999</td>
<td>504</td>
<td>grefloxacin, clarithromycin</td>
<td>10 1-3 after EOT</td>
<td>95 (EP) and 88 (ITT) 92 (EP) and 85 (ITT)</td>
<td>28-35 after EOT 90 (EP) and 75 (ITT) 89 (EP) and 76 (ITT)</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>1999</td>
<td>151</td>
<td>cefepime, ceftazidime</td>
<td>3-14 At EOT</td>
<td>79 (ITT) 75 (ITT)</td>
<td>10-14 after EOT 92 (ITT)</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>1998</td>
<td>115</td>
<td>cefepime, ceftazidime</td>
<td>3-14 3-14</td>
<td>95 (PP) and 83 (ITT) 98 (PP) and 89 (ITT)</td>
<td>Day 39-41 89 (PP) and 76 (ITT) 84 (PP) and 77 (ITT)</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>1998</td>
<td>329</td>
<td>sparfloxacin, amoxicillin</td>
<td>M 10.8 10.8 M</td>
<td>92 (PP) and 84 (ITT) 87 (PP) and 85 (ITT)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>1997</td>
<td>590</td>
<td>levofloxacin, ceftriaxone or cefuroxime, ceftriaxone</td>
<td>7-14 7-14</td>
<td>96 (PP) 90 (PP)</td>
<td>21-28 after EOT 21-28 after EOT</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>1997</td>
<td>264</td>
<td>gretloxacin, amoxicillin</td>
<td>7-10 7-10</td>
<td>76 (PP) 77 (PP)</td>
<td>28-42 after EOT 28-42 after EOT</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>1997</td>
<td>330</td>
<td>sparfloxacin, cefaclor</td>
<td>10 10 10 Day 20 + 3 10 + 3 after EOT</td>
<td>87 (PP) and 71 (ITT) 84 (PP) and 71 (ITT) 38 +/- 7 38 +/- 7 after EOT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>1997</td>
<td>690</td>
<td>cefdinir, cefaclor</td>
<td>10 6-14 days after EOT</td>
<td>89 (PP) 86 (PP)</td>
<td>21-35 after EOT 21-35 after EOT</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>1996</td>
<td>304</td>
<td>sparfloxacin, roxithromycin</td>
<td>10-14 10-14</td>
<td>95 (EP) and 84 (ITT) 80 (EP) and 75 (ITT)</td>
<td>42 after EOT 94 (EP) and 82 (ITT) 79 (EP) and 72 (ITT)</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>1996</td>
<td>1108</td>
<td>dirithromycin, erythromycin</td>
<td>10-14 10-14</td>
<td>87 (ITT) 88 (ITT)</td>
<td>7-14 after EOT 97 (ITT) 96 (ITT)</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>1996</td>
<td>298</td>
<td>ofloxacin, b-lactam with or without clavulanate</td>
<td>14.5 14.5</td>
<td>92 (EP) and 73 (ITT) 87 (EP) and 69 (ITT)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>1996</td>
<td>213</td>
<td>sparfloxacin, amoxicillin + ofloxacin</td>
<td>10 10</td>
<td>92 (PP) and 80 (ITT) 82 (PP) and 74 (ITT)</td>
<td>Day 40 + 7 87 (PP) and 69 (ITT) 78 (PP) and 69 (ITT)</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>1996</td>
<td>1137</td>
<td>sparfloxacin, amoxicillin-clavulanate or erythromycin, or amoxicillin</td>
<td>7-14 7-14</td>
<td>87 (PP) and 81 (ITT) 80 (PP) and 77 (ITT) 85 (PP) and 74 (ITT)</td>
<td>Day 42 84 (PP) and 71 (ITT) 75 (PP) and 65 (ITT) 83 (PP) and 62 (ITT)</td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>1995</td>
<td>808</td>
<td>sparfloxacin, amoxicillin-clavulanate, erythromycin</td>
<td>7-14 3-4</td>
<td>87 (PP) and 81 (ITT) 80 (PP) and 77 (ITT) 85 (PP) and 74 (ITT)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>1995</td>
<td>300</td>
<td>cefotidine, cefdinir, ceftriaxone</td>
<td>5-9 5-9</td>
<td>Good: 95 (PP) 93 (PP) 87 (PP)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* = Test of Cure, † = end of treatment, ‡ = evaluable population or per protocol
Parameters used for assessing clinical response

The criteria used to assess clinical response are shown in figure 1. Almost all studies (26/28, 93%) included signs and symptoms in their assessment of clinical response. The remaining two studies 10,11 have categorized their study patients into a cure and a failure group, but did not describe how they carried out the evaluation. Sixteen studies (57%) 10-25 incorporated radiographic chest abnormalities in the assessment of clinical response. Death by the primary diagnosis was mentioned in 9 of the 28 studies (32%) 11;16;24;26-31 as one of the criteria for failure. Of the 19 studies that did not incorporate death by the primary diagnosis, 11 studies 9;12;14;17;19;20;23;25;32-34 have described in a separate section whether the occurred deaths were associated with the primary diagnosis. Although 15 of the 28 studies 9;11;12;14;16;19;21;23;24;26;28;29;31-33 have described how many patients discontinued therapy because of adverse events, only two studies (7%) 15;26 included discontinuation of therapy as a feature of failure in assessing clinical response. 18 of the 28 studies (64%) 14-20;23;26-31;33-36 have specified an indeterminate group. Six studies 14;20;23;28;32;36 defined a recurrence group: recurrence defined as worsening of signs and symptoms after initial clinical success. If we combine recurrence and ‘relapse’, which is strictly spoken a microbiological outcome, we can conclude that 14 studies (50%) have included this parameter in their assessment of clinical response. Several studies have included other criteria in the definition for clinical response next to those mentioned by Chow and colleagues. Eleven studies (39%) 6;8;24;26;33-36 included as a criterion for failure the use of an additional or alternative antibiotic during treatment with the study drug. Four studies (14%) 12;13;19;22 involved temperature in the assessment of clinical response and three of these studies have used this parameter in addition to the parameter signs and symptoms. One study 14 also used another microbiological outcome, superinfection, in the assessment of clinical response.

Signs and Symptoms

When we further analyzed how ‘signs and symptoms’ have been used to assess cure and failure in the different studies, we observed interesting dissimilarities, as shown in fig 2. Most studies defined cure and improvement as resolution (79%) or reduction (71%) in the severity of signs and symptoms. Three studies (11%) 16;30;33 specifically mentioned reduction in the number of signs and symptoms as feature for cure, while three studies (11%) 13;18;30 defined cure as a return to baseline or pre-infection state. Most studies (75%) defined failure as persistence or no resolution of signs and symptoms. Twelve studies (43%) 13;14;16;18;23-26;31-33;36 mentioned worsening of signs and symptoms as a criterion for failure. A few studies (21%) 14;16;23;24;31-33 mentioned the development of new signs as a criterion for failure, and two studies (7%) 14;23 the lack of reduction in the majority of signs and symptoms. Patients who failed to show improvement in ≥ three clinical findings were categorized as clinical failures in one study 16.

Only 4 studies (4/28; 14%) 12;28;33;34 described what they mean by ‘signs and symptoms’. In one study 34 ‘signs and symptoms’ are described as cough, dyspnoea, crackles, rhonchi and pleuritic chest pain. In another study 33 ‘signs and symptoms’ are described as at least three of the following: cough, pyrexia, dyspnoea, decreased air entry and/or localized crackles. In two studies 12;28 also other signs and symptoms are mentioned such as absence of sputum production in the follow-up period or haemoptysis. None of the four studies described how these signs and symptoms were rated by the researchers to assign patients to a cure or failure group.
Figure 1  Parameters of clinical response used in clinical trials with community-acquired pneumonia
The first five parameters were suggested by Chow¹. The remaining three parameters (recurrence, new antibiotics needed and temperature) have been used as part of a clinical response assessment in several trials. CAP, community-acquired pneumonia, * AB = antibiotics.

Figure 2  Description of the parameters 'signs and symptoms' in the assessment of clinical cure
Time of assessment of treatment of evaluation

The time point chosen for the post treatment evaluation, indicated as Test of Cure (TOC), varies from immediately after antibiotic treatment to 21 days after treatment. Two studies have not mentioned the moment of evaluation in their results. The moment of evaluation of the late follow-up also varies between studies, with some studies (29%) not including a late follow-up evaluation moment. Of the studies that did, four studies (14%) have neither assessed a clinical response at the end of follow-up nor reported recurrence rates.

Discussion

Our systematic search of published randomized clinical trials on the antibiotic treatment of community-acquired pneumonia demonstrated that overall the parameters used for assessment of clinical response vary widely across the selected studies. The criteria as proposed by the IDSA guideline have not been systematically used. Several RCT’s have even invented their own criteria to measure clinical response, such as the presence of fever or the need for an alternative antibiotic treatment during the study period. The data convincingly show that the RCT’s use different parameters for measuring the same outcome, i.e. clinical effectiveness. This lack of standardization makes cross-study comparison difficult. Cross-study comparison is useful to study the influence of demographic characteristics such as site of care of study patients, age, co-morbid disease, microbial etiology, and immune status of the host, which, in addition to appropriate antimicrobial therapy, are likely to account for different clinical response rates.

In addition, and perhaps even more important, most parameters used were based on subjective interpretation, and are therefore prone to interobserver variability. Although the same outcome measure is applied to both the study and comparison arms of RCT’s, the use of unvalidated outcome measures might very well compromise the within-study comparisons. In most studies signs and symptoms seem to have a decisive role in the measurement of clinical response. The ability of physicians to predict pneumonia on the basis of clinical findings alone (symptoms and physical examination) has been proven to be poor, and the degree of interobserver agreement is highly variable for various physical examination findings. It is to be expected that the degree of interobserver variability is not less during recovery from the disease. Therefore, use of unvalidated signs and symptoms is not appropriate for a reliable assessment of clinical response in clinical trials. Only a few studies have included the results of the chest radiograph in the assessment of clinical response. An unresolved chest radiograph at follow-up is more likely due to underlying lung disease than to treatment failure. However, the chest radiograph can be a valuable tool in case of worsening of CAP after initial improvement due to a complication such as an empyema or the development of a new pneumonia. Although death by the primary diagnosis is a clear indication for failure, remarkably only a few studies have included this parameter as a criterion for failure. The difficulties of labelling deaths as pneumonia-related or pneumonia-unrelated are understandable as validated criteria are lacking. Mortensen et al demonstrated that pneumonia-related deaths were 7.7 times more likely to occur within 30 days of presentation compared with pneumonia-unrelated deaths. Between 30 and 90 days death was more likely to be unrelated to pneumonia. The strategy suggested by these investigators to differentiate between pneumonia-related and unrelated
mortality is to use a shorter follow-up (<30 days), or to use a clinical review committee to assign the role of CAP in the processes leading to death\textsuperscript{43}.

Several studies have included the use of additional of alternative antibiotics during treatment with the study medication as a criterion for failure. We believe that this is dangerous, because in this case investigators depend on decisions of the treating clinician for the assessment of clinical response instead of using their own standardized criteria.

What are the alternatives? As RCT's are the gold standard for evaluations of treatment effectiveness, their shortcomings in the assessment of clinical outcome should be recognised. Other outcome measures next to clinical effectiveness can also be relevant in CAP-trials, such as costs and adverse events\textsuperscript{44}. Nonetheless, when comparing new agents or new strategies, clinical effectiveness will be the single most important outcome measure.

As signs and symptoms are a key component in the assessment of clinical response, we suggest that standardized and validated symptom scores should be used to establish clinical response. In addition to the more general SF-36 scoring system, two pneumonia-specific instruments are available, the CAP-sym\textsuperscript{45} and the CAP-score\textsuperscript{46}, both validated patient-based outcome measures. The CAP-sym is developed and validated to evaluate the course of pneumonia in outpatients and measures 18 CAP-related symptoms. The CAP-score is an 8-item questionnaire, based on respiratory symptoms and on wellbeing symptoms, and fully validated for psychometric properties (i.e. reliability, validity, acceptability and responsiveness) in hospitalized patients with mild to moderate-severe CAP. Both CAP-scores are quick and easy to administer. Use of these instruments as outcome measures will allow a more objective evaluation of possible differences between treatment arms. As the CAP-scores are patient-based outcome measures, they should not be used as the only primary endpoint in controlled studies.

Other parameters which are relatively less susceptible to subjective interpretation are: all-cause 30-day mortality or death by the primary diagnosis, and discontinuation of study treatment because of side-effects or because of the individual worsening of the CAP-score, worsening of radiographic chest abnormalities or the development of new abnormalities after initial improvement. To rule out the subjectivity even more, the assignment of patients to these categories can be done by an independent clinical review committee. Finally, clinical effectiveness in RCT's is usually reported as dichotomous measure by categorizing individuals in cure and failure groups. This method is based on clinical definitions, and thus prone to interobserver variability. An alternative that is less susceptible to subjectivity in a double-blind trial is to assume cure if a patient completed follow-up without further therapeutic intervention. All other cases should be classified as failure or indeterminate according to the suggested primary endpoints.

We also noticed that the moment of TOC and the late evaluation shows great variability across RCT's. In a draft guidance of the Food and Drug Administration (FDA), released in 1998 for comment purposes (http://www.fda.gov/cder/guidance/2570dft), it is stated that the TOC visit should take place at least 7 days after the completion of treatment, assuming that the study drugs have a short half-life. The visit should take place no longer than three weeks from completion of treatment, unless the drug under study has a particularly long half-life. This is in contradiction with the Chow-criteria, which state that the clinical response should be assessed 72 hours after the end of treatment. In the draft guidance of the FDA the appropriate moment for a late evaluation is not defined, the Chow criteria suggest an assessment between 28 and 42 days.

At present, the most appropriate time of assessment of TOC and of the late evaluation moment is not clear. An assessment in the clinical setting to explore at what time therapy changes

\textsuperscript{113}
occur due to treatment failure, and at what time recurrences occur may help to identify a more rational moment of evaluation of both the TOC and the late follow-up in a study setting.

The outcome measures we propose are validated (the symptom scores), or at least objective and reproducible, in contrast to the outcome measures used in the studies we describe. The next step is consensus among trialists to standardize objective outcome measures, to allow cross-study comparisons. Certainly in this era, in which many of studies on CAP are initiated and/or sponsored by the pharmaceutical industry, it is necessary to agree upon more uniform and validated outcome measures, in order to be able to judge them properly.
Large Variability in Outcome Measures in CAP – a Systematic Review

Reference List


(2) BTS Guidelines for the Management of Community Acquired Pneumonia in Adults. Thorax 2001; 56 Suppl 4:IV1-64.


(24) Sokol WN, Jr., Sullivan JG, Acampora MD, Busman TA, Notario GF. A prospective, double-blind, multicenter study comparing clarithromycin extended-release with


Chapter 8


