Methodology for Clinical Trials Involving Patients with Cancer Who Have Febrile Neutropenia: Updated Guidelines of the Immunocompromised Host Society/Multinational Association for Supportive Care in Cancer, with Emphasis on Outpatient Studies

Ronald Feld,1 Marianne Paesmans,2 Alison G. Freifeld,3 Jean Klastersky,4 Philip A. Pizzo,5 Kenneth V. I. Rolston,6 Edward Rubenstein,7 James A. Talcott,4 and Thomas J. Walsh7

1Princess Margaret Hospital, Toronto, Ontario, Canada; 2Institute Jules Bordet, Brussels, Belgium; 3Nebraska Medical Center, Omaha, Nebraska; 4Massachusetts General Hospital, Boston, Massachusetts; 5Stanford University School of Medicine, Palo Alto, California; 6M. D. Anderson Hospital and Cancer Center, Houston, Texas; and 7National Cancer Institute, Bethesda, Maryland

Two multinational organizations, the Immunocompromised Host Society and the Multinational Association for Supportive Care in Cancer, have produced for investigators and regulatory bodies a set of guidelines on methodology for clinical trials involving patients with febrile neutropenia. The guidelines suggest that response (i.e., success of initial empirical antibiotic therapy without any modification) be determined at 72 h and again on day 5, and the reasons for modification should be stated. Blinding and stratification are to be encouraged, as should statistical consideration of trials specifically designed for showing equivalence. Patients enrolled in outpatient studies should be selected by use of a validated risk model, and patients should be carefully monitored after discharge from the hospital. Response and safety parameters should be recorded along with readmission rates. If studies use these guidelines, comparisons between studies will be simpler and will lead to further improvements in patient therapy.

Since the 1970s, many clinical trials have been conducted to compare different antibiotic regimens as initial therapy for febrile neutropenic episodes in patients with cancer. The methodology for these trials has varied tremendously, especially with regard to the issues of definition of response, stratification, trial design (with potential biases as well as other statistical problems), trial data analysis, and reporting. The methodologies used in these trials were first reviewed in 1986 by Pater and Weir [1], who looked at published reports and decided whether the studies met reasonable methodological standards on the basis of a variation of a scoring system first designed by Chalmers et al [2]. Pater and Weir [1] observed many problems, even in the “best” publications.

More recently, a consensus document published by the Immunocompromised Host Society (ICHS) in 1990 [3] recommended approaches to conducting and properly evaluating clinical trials on this subject. This consensus report served as a basis for the writing of general guidelines for the evaluation of new anti-infective drugs for the treatment of febrile episodes in neutropenic...
patients [4]. This publication was written with the financial support of the US Food and Drug Administration but was not strictly adopted by that body as formal policy [4].

Despite these publications and others in North America and in Europe [5, 6], on the basis of our review and that of others of the methodology used in published trials on this subject since the 1990s, we have observed that these recommendations have had little impact on current trial design. For this reason, a Multinational Association for Supportive Care in Cancer (MASCC) symposium was held in 1997, and 4 major issues (criteria for response, stratification, design biases, and statistical considerations) were addressed. These topics were the basis of a questionnaire that was sent to the group of experts at the symposium (including members of the executive committees of ICHS and MASCC) in an attempt to reach a consensus. Although the general conclusions were published elsewhere [7], some points will be emphasized here again.

In addition, with the growing interest in simplified therapy for some patients with febrile neutropenia (i.e., patients considered to be at low risk for developing complications during an episode of febrile neutropenia), more emphasis is warranted on trials of outpatient therapy, and it is necessary to propose recommendations specifically dedicated to their design.

**GENERAL CONSIDERATIONS**

*Definition of response to empirical antibiotic treatment and timing of assessment.* The use of a clear and standardized definition of response to therapy is needed to allow comparability between trials and the conduct of meta-analyses. The same recommendation for standardized criteria was recently published for the assessment of antitumoral response to antineoplastic treatment of solid tumors [8]. We recommend that response be defined as success of initial empirical antibiotic therapy without any modification, because this definition was used in 2 recently published articles [9, 10]. This recommendation is applicable to both hospitalized patients and outpatients. However, it is crucially important to describe the circumstances that define failure of the initial regimen, thus allowing for modifications of therapy. The analysis of trial data should include an analysis and reporting of the reasons for therapy modification and failure. The method of reporting recommendations closely depend on the type of empirical antibiotic therapy used and on the general design of the individual studies. The evaluation of response is usually performed after 72 h; we recommend that a second evaluation—one that uses the same criteria—take place on the fifth day, especially for high-risk patients, because many responding patients do not have responses until then. Time to modification is recommended as an additional potentially useful end point.

*Minimum requirements for trial data reporting.* With regard to the minimum results to be reported, persons who answered our survey recommended that the following data be included: demographic and patient characteristics, sites of infection, pathogen distribution (gram-positive, gram-negative, and the most frequently identified pathogens), resistance patterns (such as the presence of methicillin-resistant *Staphylococcus aureus*, vancomycin-resistant enterococci, or gram-negative pathogens that are resistant to extended-spectrum β-lactamases), response, modifications of initial regimen, superinfections, adverse events (possibly with the relationship to the study drugs), and deaths due to infection and due to all causes (together with time to death up to 30 days after the commencement of therapy). Data regarding trends in the neutrophil level, time to defervescence, and use of granulopoiesis-stimulating agents should be reported as well. The frequency of the administration of aggressive chemotherapy—an essential factor influencing the occurrence of febrile neu-
neutropenia—should be specified, although currently no consensus exists on the definition of an aggressive chemotherapy cycle, and no simple means exists to predict the occurrence and course of febrile neutropenia on the basis of the type of chemotherapy administered. Nevertheless, the availability of prospectively collected data might help in the future to better predict the course of febrile neutropenia according to the type of chemotherapy given. Even if they contribute some useful information, cost analyses are still not considered to be mandatory.

**Blinding.** Double-blind trials are obviously the best methodological solution, but such trials may raise ethical and logistic issues and therefore cannot always be used. Such trials should be particularly encouraged when different toxicity profiles are expected from the treatments being compared. When applicable, placebo-controlled trials—even if they are more difficult to conduct—are highly desirable, especially if intravenous therapies are being investigated. For some studies, such as the comparative evaluation of regimens such as A + B versus A + B + C, which are particularly prone to bias because of the high likelihood of the early empirical addition of the study drug (C), the placebo-controlled design is also particularly recommended. To ensure homogeneity in interpretation of the patients’ clinical courses, a data-review committee blinded to the allocated treatment is still recommended to validate all outcome assessments.

**Stratification for randomization.** Our group of experts believed that a risk-assessment model, if available and applicable, should be used as a stratification factor in randomization procedures. In addition, they recommended that the following factors also be taken into account as possible specific predictive factors related to the outcome for patients with febrile neutropenia and cancer: underlying disease as a surrogate of expected neutropenia duration that cannot be reliably assessed at the time of randomization, use of growth factors and/or antibiotic prophylaxis, and adults versus children. The treating institution is an important source of patient heterogeneity (although not specific to febrile neutropenia), and for studies that allow multiple enrollments, the type of episode (first vs. subsequent) is also strongly recommended as a factor for stratification.

**Statistical considerations.** Some general recommendations that are applicable to all randomized clinical trials were approved by our panel, such as the need for a priori sample size evaluation, a priori planning of reasons for patient exclusion from the analysis, and reporting by treatment arm of these exclusions and for performing analysis according to the principle of intention to treat. It was accepted by a majority of our experts that this intention-to-treat analysis should be performed and reported as a primary analysis, although a secondary analysis that focuses on a population of a priori defined assessable patients was judged to be extremely useful. Agreement on the need for 95% CI calculations and for adjustment of treatment comparisons for prognostic factors was also reached. A specific issue for trials conducted involving patients with febrile neutropenia is whether the same patient can be enrolled into the trial more than once. Statisticians recommend that each patient be enrolled only once; if clinicians still want to accept subsequent entries because this occurs in real life, it is recommended that a separate analysis be performed that includes only the patients’ first entries. However, subsequent entries should be accepted only if they are related to another episode of neutropenia.

Equivalence trials should be considered as a useful design for studies of febrile neutropenia, especially when response to empirical treatment is chosen as the primary end point. Indeed, at least in unselected patient populations, response rates are generally high and difficult to improve, but an equivalence trial design can be used if the experimental arm is expected to be less toxic or less costly but equally efficacious. Therefore, it might be better to think about a design aimed to show equivalence rather than choosing a difference not likely to be detected, keeping in mind than a nonsignificant result will too often be wrongly interpreted as showing equivalence between the tested treatments. This conclusion may be not adequate. A nonsignificant result does not mean equivalence but simply failure to detect the clinical difference that had been decided a priori. It then requires discussion of the choice of that difference and of the trial power. Of course, if an equivalence design is chosen, it should be clearly specified, with adequate sample size estimation and proper data analysis. The definition of equivalence should be dependent on the context and on the primary end point. It should correspond to a treatment effect less than would be a reasonable expectation in a superiority trial. We recommend that the maximum tolerated difference to consider as equivalence between treatments should not exceed 10%.

**SPECIAL CONSIDERATIONS FOR OUTPATIENT STUDIES**

It has become clear that the population of patients with febrile neutropenia seen today is highly heterogeneous in comparison with the population treated in the 1970s. The frequent use of chemotherapy for solid tumors in recent years has led to a large number of patients with less severe and less protracted neutropenia than that usually seen in patients who were treated for hematological malignancies. These patients might be at lower risk of developing complications during an episode of febrile neutropenia. The recognition of the existence of a group of patients at low risk for serious complications has had as a consequence the development and the study of new therapeutic approaches, including outpatient management strategies, that
we now want to test and compare in randomized clinical trials. However, outpatient treatment has still to be demonstrated to be as safe and as effective as hospital-based treatment in large multicenter clinical trials. Although, presently, the outpatient strategy is a successful routine practice at the M. D. Anderson Hospital and Cancer Center (Houston, Texas). The lack of real scientific demonstration of the safety of outpatient management has been pointed out by Talcott and Finberg [11]. The conduct of multicenter clinical trials of outpatient treatment raises some particular issues that are discussed below.

Selection of Patients

The 2 aforementioned recently published trials [9, 10] that compared oral treatment to intravenous treatment in low-risk patients used clinical criteria for the selection of patients to be treated in the hospital. The eligibility criteria were different in these studies. The tools used for patient selection need to evolve, because we are now trying to identify patients who are suitable for outpatient treatment. Probably the largest experience with outpatient treatment has been accumulated by the M. D. Anderson Hospital and Cancer Center Group. This group selects patients on the basis of the following criteria: patients need to be able to go home, patients need to live ≤30 miles from the cancer center, and patients must not present with a significant comorbidity (e.g., systolic blood pressure of <90 mm Hg, uncontrolled hypercalcemia, altered sensorium, and respiratory rate of >30 breaths/min or serum sodium level of <128 mM) [12]. We think that the use of a validated clinical prediction rule derived from a risk-assessment model with known discriminant parameters as well as reproducibility and transportability [13] is highly desirable, but doubt persists whether such a rule is possible to formulate.

Several models in the literature assess the risk of complications during an episode of febrile neutropenia. In adults, 2 models have so far been validated, each in studies involving several hundred patients; this is the case with the models of Talcott et al. [14] and MASCC [15]. These models are summarized in tables 1 and 2, and their respective merits are shown in table 3. However, we have to acknowledge that these models need to be regularly updated and that we still need to identify the patients in whom a low risk of complication (as predicted by the models that were largely validated for inpatients) translates into suitability for outpatient treatment. Furthermore, the currently available models do not incorporate duration of neutropenia (used in Freifeld et al. [9] and Kern et al. [10]), which is an important predictive factor, at least for response to empirical treatment, but which has not been available for identification of low-risk patients. The reason is that the value of this variable is unknown until the episode is complete. In the single-center trial reported by Freifeld et al. [9], a surrogate for duration of neutropenia—the expected duration of neutropenia determined on the basis of the chemotherapy protocols that were in use—was one criterion for enrollment in an inpatient trial of oral antibiotics but was not formally compared with other risk factors. Other surrogate factors for the duration of neutropenia may include a short period from the beginning of chemotherapy to the onset of fever and neutropenia [16], presence of underlying disease, and intensity of the chemotherapy regimen. However, we currently do not have a validated method to categorize chemotherapy regimens according to the duration of neutropenia (short, intermediate, and long).

The issue is different for children. There is one simple model [17] based solely on the monocyte count at presentation, which was prospectively derived from 227 episodes and validated by 136 episodes. This model identifies patients at low risk of developing what the authors called a “significant bacterial infection” (i.e., blood or urine culture positive for bacteria, consolidation noted on a chest radiograph, or death due to infection). Low-risk patients may be suitable for safe outpatient treatment.

<table>
<thead>
<tr>
<th>Group</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Patients hospitalized at onset of the febrile neutropenic episode</td>
</tr>
<tr>
<td>2</td>
<td>Outpatients with severe comorbidity that itself requires hospitalization</td>
</tr>
<tr>
<td>3</td>
<td>Outpatients with progressing neoplasia but without severe comorbidity</td>
</tr>
<tr>
<td>4*</td>
<td>All other patients</td>
</tr>
</tbody>
</table>

NOTE. From [14].

* Patients in group 4 are considered to be at low risk for serious complications.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Point value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burden of illness</td>
<td></td>
</tr>
<tr>
<td>No or mild symptoms</td>
<td>5</td>
</tr>
<tr>
<td>Moderate symptoms</td>
<td>3</td>
</tr>
<tr>
<td>No hypotension</td>
<td>5</td>
</tr>
<tr>
<td>No chronic obstructive pulmonary disease</td>
<td>4</td>
</tr>
<tr>
<td>Solid tumor or no previous infection in patients with hematological tumor</td>
<td>4</td>
</tr>
<tr>
<td>Outpatient status</td>
<td>3</td>
</tr>
<tr>
<td>No dehydration</td>
<td>3</td>
</tr>
<tr>
<td>Age of &lt;60 years</td>
<td>2</td>
</tr>
</tbody>
</table>

NOTE. From [15]. For each of the characteristics that is present at the onset of the febrile episode, the patients receive the corresponding number of points. The overall score is obtained by adding together the individual points (points for burden of illness are not cumulative). The maximum score is 26 points. According to the rule, patients with a score of ≥21 are considered to be at low risk for serious complication.
treatment. In the validation set, 42% of patients were identified as having low risk, with the rate of patients who were falsely identified as having low risk being 12%. However, this model should be further validated, because the 95% CI for the estimate of the rate of patients correctly identified as having low risk is presently large (95% CI, 71%–100%).

To develop outpatient therapeutic strategies, the adequate selection of low-risk patients is essential. Therefore, we suggest that the described models should be used with the goal to further validate them, under clinical conditions, as a tool for identifying patients who are eligible for outpatient treatment trials while keeping in mind that these models will need to be regularly retested and updated. Besides the use of rules based on risk quantification, we strongly recommend that psychosocial factors also be taken into account, although they are not directly related to risk.

### Monitoring Patients

An adequate patient monitoring system is a key issue in trials of outpatient therapy. Indeed, poor monitoring, even if the correct patients (and, we hope, patients who represent a high proportion of the total population of febrile neutropenic patients for optimal quality-of-life benefits) are enrolled, can lead to failure of outpatient treatment. Some minimum monitoring requirements, which are related to hospitalization length and frequency of consultations once the patient is back at home, have to be followed. Below, we propose what we consider to be some reasonable approaches.

#### Length of hospitalization

Patients are selected on the basis of an initial risk assessment that is not totally accurate and should then be followed by other assessments during hospitalization. As decided by our consensus group, the recommended length of hospitalization before the patient is discharged is 24 h, regardless of whether the patient receives intravenous or oral therapy. It may be modulated according to the time of day of the febrile episode occurrence, but an absolute minimum suggested length is 8 h, as was used successfully by Rubenstein et al. [12]. A patient’s risk status may change from low to high during the observation period in the hospital; if this occurs, the patient should not be discharged but instead should remain in the hospital to receive antibiotic therapy.

---

**Criteria for discharge from the hospital.** Early discharge from the hospital should require the patient’s acceptance, a rapid (<2 h) access to a medical facility that is familiar with the modern outpatient management of febrile neutropenia, and phone availability. For potentially better compliance with treatment and safety preventative measures, especially when oral therapy is given, it is recommended to send home only those patients who have dedicated and effective household care.

**Monitoring after discharge from the hospital.** Patients should be examined, at home or in the hospital, by a nurse, a general practitioner, or a specialist physician each day during the 3 first days after discharge and then every 2 days until resolution of fever and neutropenia. Blood tests should be performed every 2 days. Although it is difficult to perform, assessment of patient compliance with treatment should be undertaken, but with careful interpretation of its impact on efficacy. Patients who were compliant with placebo regimens might have better outcomes than do noncompliant patients [18].

These criteria are not proven requirements for care. However, for the time being, we think that noncompliance with these recommendations could be unsafe for the patients, particularly in multicenter clinical trials.

### Outcomes

As already mentioned, the primary outcome of such studies should be the same as that used for studies involving hospitalized patients—namely, the response rate to initial antibiotic therapy without modification. A careful and complete description of the reasons for change is also required. An important secondary outcome is the need for rehospitalization. However, all readmissions should not be considered indicative of failure of empirical treatment, and the reasons for readmission to the hospital should be carefully reported and classified according to the following categories: admission because of infection, admission because of toxicity, or admission for another reason. Particular attention should be paid to any adverse effect, even if it is minor. The reason for keeping a patient as an inpatient should be assessed by means of the same categorization. For the studies that compare inpatient management with outpatient management (early discharge), an outcome directly related to the study hypothesis is appropriate; in this situation, designs for showing equivalence are considered to be appropriate. The primary remaining question about outpatient management of febrile neutropenia is its safety relative to conventional inpatient care. Safety can be assessed by use of the occurrence of major

---

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Talcott et al. [14]</td>
<td>MASCC(^a) [15]</td>
</tr>
<tr>
<td>Patients at low risk</td>
<td>26 (21–30)</td>
<td>63 (59–68)</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>93 (88–98)</td>
<td>91 (87–94)</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>23 (18–28)</td>
<td>36 (28–44)</td>
</tr>
<tr>
<td>Specificity</td>
<td>90 (84–97)</td>
<td>68 (58–79)</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>30 (25–35)</td>
<td>71 (66–76)</td>
</tr>
<tr>
<td>Global misclassification rate</td>
<td>59 (54–64)</td>
<td>30 (25–34)</td>
</tr>
<tr>
<td>No. of deaths among low-risk patients</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

**NOTE.** Data are % (95% CI), unless otherwise indicated.

\(^a\) For the validation set, \(n = 383\).
medical complications as an outcome measure [14–16], including potentially life-threatening or lethal events.

Finally, because outpatient therapy is designed to improve the quality of life of the patients as well as to decrease the cost of the treatment of febrile neutropenia, some evaluation of these parameters is highly recommended in future trials that evaluate outpatient management of febrile neutropenia. Although cost-effectiveness can be reasonably assumed for safe outpatient treatment strategies compared with inpatient management, the price of a hospital day is a very important component. Most hospitalizations for febrile neutropenia for low-risk patients are 5–10 days in duration, and this should cost less than provision of the same therapy at home. However, few hard data exist on this subject, and the costs of home monitoring, patient visits to the hospital, or caregiver visits to the patient at home are considerable. Possible patient readmissions may significantly reduce the cost savings associated with outpatient therapy. We look forward to future studies that evaluate this issue.

CONCLUSION

Ideally, we recommend that all future trials conducted that involve patients with febrile neutropenia (either inpatients or outpatients) should be performed in a manner consistent with these guidelines. It would be extremely helpful if regulatory agencies worldwide could approve these or similar guidelines and, in the future, require pharmaceutical companies to use them to design and perform trials with new antimicrobial drugs to get their product accepted for the therapeutic indication of “febrile neutropenia.”

Acknowledgments

We thank Dr. Linda Elting (M. D. Anderson Hospital and Cancer Center, Houston, Texas) for her help with the development of the initial questionnaire. We thank Aileen Urchuk for her assistance with the preparation of the article in manuscript.

References