Diabetic foot disease and particularly foot ulceration and amputation, result in a major global burden for patients and the health care system and represent a critical source of suffering and high financial cost for patients, family members, care givers, heath care professional and facilities and society in general. The global prevalence of diabetic foot varies from 3% in Oceania to 13% in North America, with a global average of 6.4%. Some authors consider that diabetic foot ulcers have been neglected in health-care research and planning, and clinical practice is based more on opinion than scientific fact. Robust clinical trials of management, using appropriate end-points are welcome to be implemented.

According to a recent systematic review, the assessment and comparison of published trials remains difficult with marked clinical heterogeneity between studies. Despite the publication in 2016 of the 21-point assessment criteria developed jointly by the IWGDF and the European Wound Management Association to guide improvements in trial design and presentation of study findings, the majority of the new articles reviewed was still at a moderate to high risk of bias. The overall conclusion has been that with very few exceptions, the evidence available from published studies is of insufficient quality to recommend any particular treatment or dressing product in preference to any other. Therefore, a good adherence to preventive guidelines are mandatory.

One of the challenges in this particular disease is the complexity of comparing one patient to another due to the fact that crucial variables are not always clearly disclosed by the authors, including duration of metabolic disease, previous episodes due its repetitive nature, amputation history, concomitant renal or retinal disease, degree of metabolic decompensation, diabetic peripheral neuropathy and peripheral artery disease grades, infection association and wound duration, which usually are fundamental to understand the differences in the magnitude of response.

Most of the researchers agree that the introduction of treatments that lack substantial evidence of effectiveness should be avoided. However, many trials, particularly those with small samples sizes claim to have found effective measures to treat such a complex disease, due to a high effect size despite a probability of bias during planning and conduction of the study. The increase in sample sizes is generally viewed as a favorable development. Larger sample sizes provide more power to identify a treatment effect that is really present. In addition, the effect is
estimated more precisely (smaller confidence intervals). Larger trials are also a natural consequence of head-to-head trials because the difference between two active drugs is generally expected to be small, and therefore, the required sample size needs to be relatively high. However, larger sample sizes also make trials expensive and time consuming. This can be a barrier for non-commercial investigators to perform a trial. Moreover, it can be ethically questionable to ask more patients to participate, especially when the safety of the tested drug has not yet been established. Another disadvantage of (very) large sample size is that a difference in outcomes between the groups will become (very) statistically significant, no matter how small or clinically meaningless it is. Additionally, up to 52% of published studies show discrepancies for sample size calculations between protocols and publications. Therefore, some experts advise that focusing on the magnitude of response, such as the effect size, may be more relevant than in the size of the study population.

Table 1 shows the results of a small-sample and two large-sample studies in the field of diabetic foot ulcer treatment. Common statistical values are depicted, including effect sizes in experimental and control groups, points of difference, alpha, post hoc beta power and the number needed to treat (NNT). As can be shown small-sample studies require a lower NNT than large-sample studies, indicating that more patients will be needed to find at least a new patient responding to the experimental intervention. It is also remarkable that post hoc beta value, a measure of the power of the study is greater in the first small-sample study.

Clinicians commonly judge results from clinical trials based on their interpretation of P values, a measure of statistical significance only, without even considering the effect size. While a significant P value suggests that something nonrandom has occurred, it does not inform us about the clinical significance of the nonrandom effect. For example, the magnitude of statistical significance is heavily influenced by the number of patients studied. Other things being equal, larger samples yield more significant P values than smaller samples. Consequently, a large trial of a marginally effective treatment can have greater statistical significance than a small trial of a highly effective treatment.

On the contrary, effect sizes inform clinicians about the magnitude of treatment effects. A major dilemma is the interpretation of an effect size which still requires evaluation of the meaningfulness of the clinical change and consideration of the study size and the variability of the results. For non-continuous measurements or for binary (success/failure in diabetic foot ulcer cure) outcomes, the best method of calculating the effect size is the NNT. The NNT is defined as the number of subjects one would expect to treat with agent A to have one more success (or one less failure) than if the same number were treated with agent B. NNT is a measure related to absolute risk reduction and may be most useful in assessing relevance of treatment effects.
The lowest NNT the better magnitude of treatment effect. The highest NNT the lowest power of the proposed effect. It should be noted that in the small study by Janka-Zires, the NNT was only 2.6, compared to 5.6 and 8.3 found in the larger studies by Edmons and Gant. Therefore, the evaluation of the magnitude of response appears to be what really matters.

In conclusion, we should open our minds to focus primarily on the effect size rather than the sample size when we need to assess the success of a new experimental intervention.
References:


Table 1. Main statistical findings in small and large-size study examples

<table>
<thead>
<tr>
<th>Principal investigator</th>
<th>Year</th>
<th>Intervention, duration</th>
<th>Effect size (%) experimental Group (Sample size)</th>
<th>Effect size (%) control group (Sample size)</th>
<th>Points of Difference</th>
<th>Alpha power</th>
<th>P value</th>
<th>1 - Beta power</th>
<th>Post Hoc</th>
<th>NNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Janka-Zires M</td>
<td>2016</td>
<td>Topical 8% Pirfenidone gel Conventional treatment 8 weeks</td>
<td>52.4 (n=21)</td>
<td>14.3 (n=14)</td>
<td>38.1</td>
<td>0.024</td>
<td>0.65</td>
<td>2.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edmons M</td>
<td>2017</td>
<td>Sucrose octasulfate dressing placebo dressing Good Standard of Care 20 weeks</td>
<td>48.0 (n=108)</td>
<td>30.0 (n=95)</td>
<td>18</td>
<td>0.0045</td>
<td>0.60</td>
<td>5.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gant F</td>
<td>2018</td>
<td>Leucocyte, fibrin and platelet patch versus Good Standard of care 20 weeks</td>
<td>34.0 (n=132)</td>
<td>22.0 (n=137)</td>
<td>12</td>
<td>0.0235</td>
<td>0.58</td>
<td>8.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Beta = probability of type II error and 1 - B = power test according to G power software in post hoc analysis.
NNT = number needed to treat, NNT = 100 / (% improved on treatment - % improved on placebo).