Readers should ask themselves two questions when critically reading the article by Waagfjord et al: 1) How convincing is the authors' evidence that the intervention (treadmill walking) actually caused a positive change in the base of support and symmetry of the subject's gait and no change in the remaining variables? and 2) Does the reader believe that the change in gait variables represents a meaningful change in the subject's function? I will address these questions separately.

The first question deals with the internal validity of the experiment. Threats to internal validity translate to the reader as "What else could have caused the changes in the gait measurements other than the authors' claim that the changes were caused by their intervention?" The reader must examine the data and be convinced that any of the common threats to internal validity are not plausible explanations for the authors' results. I will attempt to pick out various portions of their data and point out other possible causes for the trends that were seen.

In examining the graphical presentation of the data in Figures 2, 5, 6, and 7, the reader is struck by an obvious upward trend in the data during the baseline phase. This trend was clearly not expected by the authors, given their criteria for subject selection ("the recovery interval was selected ... to minimize functional gains independent of the intervention"). Thus, a stable, level trend was expected during the baseline period, yet, a clear upward trend was obtained.

The possible conclusions drawn from this upward trend are very much the same with this single-subject experiment as they would be with any research design. For example, one interpretation of the upward trend in the data during the baseline phase is that the dependent measure(s) are simply unreliable in the sense that they lack agreement. This statement is made despite the fact that acceptable reliability coefficients are offered by the authors. Closer examination of the basis of the reliability coefficients, however, render the authors' conclusion of acceptable reliability misleading. The principal investigator measured the footprint data twice, and the paired measurements were subjected to a reliability estimate. Thus, the readers are offered a reliability coefficient based on the investigator's ability to use a ruler to measure distances between ink blots. But what about the repeatability of the subject's performance, with agreement being made during the baseline period, yet, a clear upward trend was obtained.

This upward trend could also be interpreted as being caused by any of the well-recognized threats to internal validity.1 For example, with the effect of testing, the expected change (eg, increased velocity because of the treadmill exercise) is caused by the number of times a particular response (the dependent measure) was measured before the intervention rather than by the intervention itself. The effect of testing can arguably be a greater problem when the dependent measure is obtained repeatedly, such as it is with any single-case research design.

Whatever the reason, the upward trend in the data presents a problem when interpreting the results, particularly when attributing cause to the intervention. It appears to me that for some variables, greater effects occurred during the baseline phase than during the intervention phase of the experiment. Some of the data suggest that the intervention actually interrupts some of the "progress" being made during the baseline phase (Figs. 2, 5, 6, 7, 8). This statement is based on my own visual analysis. This issue illustrates a major problem with visual analysis of single-case research data, namely that the technique of visual analysis lacks any formal decision rules to guide the interpretation of the data. Thus, I can interpret the data for a variable such as velocity as...
a possible negative effect of the intervention, and the authors can interpret it totally differently as "no effect."

An alternative to visual analysis is to statistically analyze single-case experimental data. Statistical analysis solves the major problem of visual analysis, but is not without problems itself. One of the major issues with statistical analysis is the problem of serial dependency in the data, which the authors point out. For example, with selected variables with upward trends and concomitant significant autocorrelation (eg, velocity), the authors chose not to further analyze their data using the C statistic and instead used visual analysis. When faced with significant trends in the baseline data, however, Ottenbacher suggests that a celeration line technique be used in conjunction with the C statistic as an alternative analysis procedure. In addition, there are other procedures that can be used to statistically analyze single-case research data using celeration line techniques. Admittedly, some alternative techniques do have limitations, namely that the Type I error rate will be inflated. The C statistic, however, has been criticized for the same reason. I am curious as to why the authors chose to rely on visual analysis for selected variables (eg, velocity) that exhibited autocorrelation and not other variables that also exhibited significant autocorrelation (eg, treatment withdrawal for right step length).

In terms of the base-of-support findings, there are a few points that I wish to raise, followed by an alternative explanation to the authors' findings. First, an alternative analysis can be performed by extending the celeration line from the baseline phase into the treatment phase and calculating a binomial statistic. This analysis shows that two points are above the extended celeration line and five points are below it (P = .0703). Thus, an alternative analysis does not support the authors' contention of an intervention effect. This interpretation is further strengthened given that the limitation of the binomial testing is a possible inflation of the Type I error rate. Second, the C statistic is best used to quantify the significance of trends in the data. Trends can be represented by differences in the level of the celeration lines. Trends can also be represented by changes in slope of the celeration lines of the baseline and intervention phases. Some would argue that they are best represented by slopes of the celeration lines (Ottenbacher KJ; personal communication, June 5, 1990). Looking at the graph pertaining to the base-of-support data (Fig. 4), it appears that the celeration line of the baseline phase is decreasing (a desired effect) and that the slope of the celeration line in the intervention phase is increasing (an undesired effect). Instead of noting the rather small difference in levels of the celeration lines and assuming the data support a decreasing trend in base of support, another equally plausible explanation would be that the C statistic is significant for an increasing trend in base of support during the intervention, based on the slopes of the celeration lines.

A similar, but stronger, argument could be made for an alternative explanation to the authors' interpretation of the right step-length data. A split-middle technique using a binomial test would again be nonsignificant for any change in right step-length trend during the intervention. The significant C statistic could be argued to support a decreasing rather than increasing trend because of the increased celeration line slope during the baseline phase and the decreasing slope during the intervention phase. The slope difference is even more marked in the right step length than the base-of-support data. Furthermore, the level of the celeration line during the intervention phase is less than the level during the baseline phase. Taken together, it would seem that these issues would support any significant intervention effect being a decreasing rather than an increasing trend.

My final point deals with the meaningfulness of the change in base-of-support measurements seen in patients. The authors admit that the change in the base of support was small. With many clinical measures in physical therapy, it is difficult to visualize exactly how large a change is necessary before one can make the assertion that a "clinically meaningful change" has occurred. I think this difficulty is compounded with measures that are not routinely quantified, such as the base-of-support measure. The authors make a good point in relating the small change in base-of-support measurements to possibly meaningful decreases in energy expenditure during everyday walking. I only wish to point out just how small this change is. Looking at the difference in levels of the celeration lines in the baseline and the intervention phases (Fig. 4), there appears to have been about a 0.95-cm change in the base-of-support measurements. Given the apparent error associated with this measure (consider the instability of the baseline measure), I wonder whether a clinician can detect a change in the base of support that is of such small magnitude.

Anthony Delitto, MHS, PT
Assistant Professor
Program in Physical Therapy
Washington University School of Medicine
660 S Euclid Ave
PO Box 8083
St Louis, MO 63110

References

4 Crosbie J. The inability of the binomial test to control Type I error with single-subject data. Behavioral Assessment 1987;9:141-150.