Factors Confounding Assessment of Ambulatory Blood Pressure Monitors, Studied During Formal Evaluation of the Tycos Quiet-Trak

J. Conor O'Shea and Michael B. Murphy

The credibility of studies assessing ambulatory blood pressure monitoring (ABPM) devices will be enhanced by minimizing opportunities to manipulate the outcome through biased selection of patient volunteers, and by insuring that participants are representative of the target patient population. While subjecting the Welch Allyn Quiet-Trak to the British Hypertension Society (BHS) protocol, we examined the extent to which certain candidate characteristics might influence device assessment.

The Quiet-Trak achieved an A grade overall. During its field testing, daytime physical activity measured with a wrist-mounted, piezoelectric accelerometer, influenced significantly the rate of measurement rejection. The tertiles of subjects with highest and lowest levels of daytime physical activity (64 ± 16 and 35 ± 10 activity units; P < .001) exhibited significantly different measurement rejection rates (10 ± 3 and 3 ± 2 daytime rejects; P < .001). Most rejected readings occurred during episodes of high physical activity. During static evaluation (Phase V), the level of systolic BP influenced the accuracy of diastolic BP estimation; above 190 mm Hg systolic BP, estimates of diastolic BP differed significantly from manual measurements. Subjects' age and arm circumference influenced neither field nor static evaluation.

Retrospective comparison of the study subjects with 120 consecutive ABPM clinic attenders revealed (1) that participants in field testing were younger, had lower BP but were equally active compared with clinic patients; and (2) that Phase V participants exhibited higher systolic (155 ± 42 vs 135 ± 15 mm Hg; P < .001) but similar diastolic BP levels compared to controls.

The reliability of ABPM validation protocols would be enhanced by: (a) incorporation of objective measurement of physical activity during field testing, demonstrating normal levels of activity; (b) requiring that field testing be conducted in representative patients rather than normal volunteers; and (c) a greater focus in both static and field testing on those levels of blood pressure that are relevant in clinical practice.

KEY WORDS: Ambulatory blood pressure monitoring, validation protocol, physical activity, Quiet-Trak.
theless, despite further revisions of both protocols in 1992 and 1993, respectively, there is still scope for improving the evaluation procedure.

Experience in clinical drug trials has shown that investigators tend to select participants carefully to ensure efficient completion of study protocols. Subjects likely to comply with the investigators' instructions may be included preferentially, whereas those who may not adhere to the protocol or who may dropout, are excluded. As a result, the study groups are often poorly representative of the population for whom the intervention was designed.

In an ABPM validation study, it may be possible to manipulate subject selection to influence performance rating or to accelerate the completion of the study, while remaining in compliance with the BHS protocol. For example, field testing of the model is an important aspect of the validation procedure. Three devices must be worn for 24 h by each of 8 subjects and it is essential that at least 70% of the BP readings are valid for 22 of the 24 recording days. With prior knowledge of the factors determining the rate of measurement rejection an investigator could enhance the likelihood of achieving acceptable data or, on the other hand, ensure failure, in the field testing of a particular device by selecting subjects with the desired characteristics. Furthermore, such selection might result in a performance assessment that would not be achieved in the population requiring ABPM in routine clinical practice.

Accordingly, in the course of evaluating the Quiet-Trak (Tycos, Welch Allyn Inc., Arden, NC; also marketed as Tenso 24) according to the BHS (1990) validation protocol, we studied the contribution of various candidate biases to the outcome of the study. First, we examined the influence of the level of subjects' physical activity, quantified electronically, on the number of rejected/aborted BP data during the 24-h field test. Second, a retrospective analysis was performed to determine the extent to which subjects' age, arm circumference, and ambulatory BP levels might have influenced the outcome of field testing. Finally, the manner in which these characteristics determined the results of Phase V accuracy testing and the extent to which study participants represented clinical practice, was also assessed. Investigators were not aware of the intention to conduct retrospective comparisons at the time the validation study was conducted.

METHODS

The Ambulatory BP Recorder  The Quiet-Trak ambulatory pressure recorder is an automated, noninvasive device developed by Tycos-Welch-Allyn Inc., Arden, NC. It is small (11.4 x 8.6 x 4.1 cm), lightweight (355 g) and uses a microphone on the inflatable cuff to detect Korotkoff sounds. The system consists of a portable recorder, powered by four alkaline AAA (1.5 V) batteries. A connector on the inflation tube allows calibration against a mercury sphygmomanometer. Data stored in the recorder can be downloaded to an IBM-compatible PC for analysis by dedicated software. Measurements can be made over the ranges 60 to 250 mm Hg for systolic blood pressure (SBP) and 30 to 160 mm Hg for diastolic blood pressure (DBP) and 40 to 160 beats/min for the heart rate. Measurement intervals may range from 5 min to 2 h. The typical cuff inflation time is 15 sec, but this varies with cuff size and target pressure. The deflation rate is adjustable from 2 to 6 mm Hg per second. It does not provide a warning beep prior to recording.

The Activity Recorder During the in-field performance testing of the Quiet-Trak, each of the 24 subjects wore an electronic activity monitor (Gaehwiler Electronics, Hombrechtikon, Switzerland) on the dominant wrist. This battery powered unit, slightly larger than a wristwatch, is a monaxial piezoelectric accelerometer, programmed to record activity every 10 sec. It is designed to integrate motor activity (threshold 0.1 g) over a defined time period, and convert it into an activity value ranging from 0 to 253. Since the sampling time of the monitor is 125 msec (8 Hz), 80 separate measurements are made for each of the 10-sec data points recorded. We have previously demonstrated that physical activity measured at the dominant wrist using this device is a significant determinant of BP variability.

Data were analyzed in 21 of the 24 subjects, with battery failure rendering two of the recordings unusable, and failure to download the data correctly excluding the third.

The British Hypertension Society Protocol Program This consisted of five phases, as originally published. These are as follows. Phase I was observer training and assessment. Phase II was before-use intermonitor variability assessment (three devices). Phase III was in-use assessment or field testing. To evaluate the clinical performance of the Tycos Quiet-Trak, the monitors used in Phase II were tested over a 24-h ambulatory recording period. The three monitors were worn by 24 volunteers for 24 h, yielding 8 records for each. BP measurements were made at 15-min intervals from 09:00 to 22:00 (56 measurements), and at 30-min intervals from 22:30 to 08:30 (19 measurements), giving a total of 75 readings for the 24 h. Phase IV was after-use intermonitor variability assessment. Phase V was static evaluation of device accuracy. In this phase, three BP measurements were made simultaneously using the Tycos Quiet-Trak unit and a mercury sphygmomanometer using a Y-connector in 85 subjects, resulting in 255 paired measurements of systolic and diastolic blood pressures. The 85 subjects were identified following review of BP records of 250 hospital patients. A total of 94 subjects were screened. Nine patients were rejected: four because of atrial fibrillation at the time of screen-
ing, two because Korotkoff sounds persisted to near zero, two because their blood pressure at the time of screening was outside the required range, and one because of repeated "Code 2" readings from the device indicating persistent air-leak. Finally, the test device was graded, based on the percentage of test device measurements that differed from the mercury standard by less than 5, 10, and 15 mm Hg, according to the protocol.

The Control Clinical Practice Population A total of 120 consecutive patients attending our ABPM unit, beginning on November 1, 1993, with simultaneously recorded activity using the Gaehwiler device, provided age, arm circumference, blood pressure, and physical activity data to establish the norms of our clinical practice.

Statistical Analysis Summary data are described as mean ± standard deviation (SD). Comparisons were made between groups by paired and unpaired t tests, as appropriate, with P < .05 accepted as statistically significant. The contributions of age, arm circumference, blood pressure levels, and the extent of physical activity to the measures of device performance were evaluated by multiple regression analysis, using the Astute Statistical Software program (DDU Software, Leeds University, England).

RESULTS

BHS Protocol Outcome Phase I: Observer Training and Assessment The subjects' pressures ranged from 110/62 to 194/102. A total of 98% of systolic and diastolic pressure estimates differed between the expert and the trainees by < 5 mm Hg and none differed by more than 10 mm Hg.

Phase II: Before-use Intermonitor Variability Assessment One hundred percent of the Quiet-Trak measurements were within 2 mm Hg of the mercury sphygmomanometer readings.

Phase III: In-use (Field) Assessment For each of the 24 recordings there were, on average, 83 inflations and 73 (88%) of these readings were valid. There were 4.3 rejects and 5.6 aborted readings per recording; there were 51.9 and 20.6 valid readings per day and night respectively giving a day/night valid reading ratio of 2.3. Twenty-three of the recordings met the criterion of >70% valid measurements (the 24th recording had 68% valid measurements), thus meeting the BHS requirements.

Patient Acceptability Twenty-three of the 24 subjects completed a questionnaire on the performance of the device. Only two of the volunteers had worn an ABPM device before. One of these expressed a preference for a previously tried device; the other choose the Tycos Quiet-Trak. Apart from the inconvenience of not being able to shower or swim, the subjects generally had few problems with the monitors. One subject developed an area of erythema at the site of the microphone placement.

Phase IV: After-use Intermonitor Variability Assessment One hundred percent of the measurements were within 2 mm Hg of the corresponding sphygmomanometer readings.

Phase V: Static Device Accuracy Validation The Quiet-Trak achieved an A grade (Table 1), complementing the previous assessment using the AAMI guidelines. The differences between the observer and device estimates of BP are plotted against the average of the two estimates (for each of the 255 measurements in the 85 subjects) in Figures 1 and 2 (systolic and diastolic BP respectively). SBP was measured accurately by the Quiet-Trak throughout the entire BP range; it underestimated SBP when compared with the observers by 1.5 ± 4 mm Hg.

The diastolic BP data were more complex. Overall, the Quiet-Trak overestimated diastolic BP compared with the observers, by 2.5 ± 18 mm Hg. However, Figure 2 illustrates that the preponderance of this diversity may be attributed to 16 highly inaccurate measurements (difference from observer’s estimate >20 mm Hg) in 7 subjects, where the device overestimated

<table>
<thead>
<tr>
<th>Observer/Device</th>
<th>n</th>
<th>Mean ± SD</th>
<th>Diff ± SD</th>
<th>% ≤5</th>
<th>% ≤10</th>
<th>% ≤15</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer 1, Systolic</td>
<td>129</td>
<td>146 ± 34</td>
<td>-11 ± 3</td>
<td>90</td>
<td>98</td>
<td>100</td>
<td>A</td>
</tr>
<tr>
<td>Tycos, Systolic</td>
<td>129</td>
<td>144 ± 34</td>
<td>-0 ± 3</td>
<td>84</td>
<td>91</td>
<td>95</td>
<td>A</td>
</tr>
<tr>
<td>Observer 1, Diastolic</td>
<td>129</td>
<td>85 ± 20</td>
<td>0 ± 10</td>
<td>84</td>
<td>91</td>
<td>95</td>
<td>A</td>
</tr>
<tr>
<td>Tycos, Diastolic</td>
<td>129</td>
<td>44 ± 25</td>
<td>-5 ± 5</td>
<td>85</td>
<td>95</td>
<td>98</td>
<td>A</td>
</tr>
<tr>
<td>Observer 2, Systolic</td>
<td>126</td>
<td>165 ± 17</td>
<td>-0 ± 3</td>
<td>71</td>
<td>88</td>
<td>91</td>
<td>C</td>
</tr>
<tr>
<td>Tycos, Systolic</td>
<td>126</td>
<td>160 ± 44</td>
<td>-5 ± 24</td>
<td>71</td>
<td>88</td>
<td>91</td>
<td>C</td>
</tr>
<tr>
<td>Observer 2, Diastolic</td>
<td>126</td>
<td>87 ± 16</td>
<td>-5 ± 24</td>
<td>71</td>
<td>88</td>
<td>91</td>
<td>C</td>
</tr>
<tr>
<td>Tycos, Diastolic</td>
<td>126</td>
<td>92 ± 31</td>
<td>-5 ± 24</td>
<td>71</td>
<td>88</td>
<td>91</td>
<td>C</td>
</tr>
<tr>
<td>Overall Grade</td>
<td></td>
<td></td>
<td></td>
<td>82</td>
<td>93</td>
<td>96</td>
<td>A</td>
</tr>
</tbody>
</table>

Observers 1 and 2 measured blood pressure in 42 and 43 subjects respectively. Diff. difference. % Readings < mm Hg.
DBP by 69 ± 25 mm Hg. These subjects were among those with the highest systolic blood pressure, their average sphygmomanometric pressure being 205 ± 25/99 ± 10 mm Hg. The average age of these subjects was 64 ± 7 years and their average arm circumference was 29 ± 3 cm. The outlying diastolic estimates were recorded by both observers.

As the measurement inaccuracy seemed to be confined to levels of systolic pressure rarely seen in ABPM practice (vide infra), we reexamined the data to compare device performance in patients with SBP less than or greater than 190 mm Hg. At higher systolic pressure levels (classified by SBP > 190 mm Hg at the first sphygmomanometric measurement; \( n = 20 \), average SBP 215 ± 15 mm Hg), the Tycos Quiet-Trak underestimated SBP by 7.1 ± 5.0 mm Hg compared with the sphygmomanometer. However, it overestimated diastolic BP by 14.0 ± 3.5 mm Hg. In subjects whose SBP was less than 190 mm Hg (\( n = 65 \); average SBP 136 ± 26 mm Hg) it underestimated SBP by 1.1 ± 5.0 mm Hg and underestimated DBP by 1.6 ± 3 mm Hg.

Factors Potentially Influencing Protocol Outcome

Field Testing (Phase III) (Table 2) Multiple linear regression analysis showed that, of the characteristics evaluated, only average awake physical activity level was a significant determinant of performance, accounting for up to 24% of measurement rejections. Subjects' age, arm circumference, and level of systolic pressure were noncontributors. When the volunteers were divided into tertiles based on the average level of daytime physical activity, the tertiles of subjects with highest and lowest levels of physical activity (64 ± 16 and 35 ± 10 activity units \( P < .001 \)) exhibited significantly different measurement rejection rates (10 ± 3 and 3 ± 2 daytime rejects; \( P < .001 \)). There was no significant difference between these tertiles in average age, arm circumference, or BP levels. Not surprisingly, rejected measurements were associated with a significantly higher activity level at the time of BP estimation (88.4 ± 36 vs 40.7 ± 17; \( P < .001 \)) than those measurements deemed acceptable.

The average activity score of the subjects who participated in the Phase III field testing (49 ± 15 units) was similar to that of the control population of 120 consecutive ABPM clinic patients (43 ± 15 units; \( P > .05 \)). Although they were also similar in arm circumference distribution, they were significantly younger and exhibited lower BP (Table 3).

### Table 2. Characteristics of the “Field Test” Subjects Ranked by the Number of Rejected Daytime BP Data

<table>
<thead>
<tr>
<th>Tertile</th>
<th>Rejected BP Data (daytime)</th>
<th>Age (years)</th>
<th>SBP (mm Hg)</th>
<th>DBP (mm Hg)</th>
<th>Arm Circ. (cm)</th>
<th>Activity Score (daytime)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest</td>
<td>10 ± 2</td>
<td>31 ± 5</td>
<td>122 ± 9</td>
<td>77 ± 6</td>
<td>26 ± 6</td>
<td>63 ± 17</td>
</tr>
<tr>
<td>Middle</td>
<td>5 ± 2</td>
<td>32 ± 5</td>
<td>113 ± 4</td>
<td>70 ± 4</td>
<td>24 ± 5</td>
<td>50 ± 2</td>
</tr>
<tr>
<td>Lowest</td>
<td>3 ± 2</td>
<td>31 ± 4</td>
<td>114 ± 9</td>
<td>73 ± 6</td>
<td>26 ± 5</td>
<td>34 ± 12</td>
</tr>
</tbody>
</table>

* \( P < .01 \); * \( P < .005 \) highest versus lowest tertile.

BP, blood pressure; Circ., circumference; DBP, diastolic blood pressure; SBP, systolic blood pressure.

![FIGURE 1. The differences between two observer and device estimates of systolic blood pressure (SBP) plotted against the average of the two estimates (for each of the 255 measurements in the 85 subjects).](image1)

![FIGURE 2. The differences between observer and device estimates of diastolic blood pressure (DBP) plotted against the average of the two estimates (for each of the 255 measurements in the 85 subjects).](image2)
Static Evaluation (Phase V)  The characteristics of the 85 participants in the validation phase are also compared with the clinic controls in Table 3. Apart from exhibiting significantly higher systolic BP (115 ± 42 mm Hg; \( P < .0001 \)), study subjects were similar to controls.

**DISCUSSION**

The objectives of the study were threefold: to evaluate the performance of the Quiet-Trak and in so doing, to scrutinize the factors that could be manipulated, if desired, to influence the outcome; and to examine, retrospectively, the extent to which participants mirrored clinical practice. In brief, the device achieved a Grade A (Table 2) during Phase V static evaluation of measurement accuracy in volunteers who were demonstrably representative, in most respects, of the usual ABPM clinic population. However, the study identified several weaknesses in the evaluation process, only some of which have been addressed in the most recent version of the protocol.

**Phase V: Static Evaluation**  Although it provided highly accurate estimates of systolic BP throughout a broad range of pressure, the Quiet-Trak performed poorly at estimating diastolic pressure when the systolic BP exceeded 190 mm Hg. This finding underlines the importance of the shift in focus of the newer guidelines from a single composite grade to the need to demonstrate device reliability in the particular patient groups for which it is intended. As defined by BP levels or demographic characteristics. On the other hand, whether the poorer performance above SBP of 190 mm Hg is clinically important is doubtful. The major role of ABPM is to aid the clinician in determining whether drug treatment is indicated in patients with “borderline” or moderate elevations in sphygmomanometric readings, and to evaluate the quality of control in treated patients. However, our clinic control data indicate that 190 mm Hg is over 3 SDS above the mean pressure of those ordinarily monitored for the purposes. In the case of the second most common indication, that of evaluating the efficacy of antihypertensive drug therapy in clinical trials, ethical considerations usually preclude patients with this degree of hypertension. This is evidenced also by scrutiny of recently published studies where 190 mm Hg is 2.5 SDS above the mean pressure of the participants. 11 Reassuringly, the most recent revision of the BHS protocol recognizes these facts, by substantially reducing the requirement for inclusion of such severely hypertensive subjects (from 30% in the SBP 180 to 240 mm Hg range,11 to eight of 85 subjects,14 concentrating instead on those BP levels most relevant in clinical practice. We cannot, of course, infer that accuracy at rest will be translated into accuracy in the ambulatory setting.

Although the device was not subjected to the AAMI protocol, the Phase V (static evaluation) results can be analyzed according to AAMI criteria. These require a difference of less than 5 ± 8 mm Hg between the test device and the mercury standard. On this basis, the Quiet-Trak recorder met the AAMI criteria for SBP but not for DBP, where the mean difference was 2.5 ± 18 mm Hg. However, in subjects with SBP < 190 mm Hg, the mean difference between the Tycos device and the mercury standard was -1.1 ± 5 mm Hg for SBP and -1.6 ± 3 mm Hg for DBP, both within the AAMI limits.

We found no evidence that demographic characteristics such as age or arm circumference determined outcome in Phase V. The absence of any influence of arm circumference in either field testing or static evaluation, which one might also intuitively expect, may be explained by the compensatory use of a large cuff when the circumference exceeded 31 cm. That age did not influence the data was surprising as Miller and others have reported increasing discrepancies, with higher subject age, between manual measurements and those recorded by an earlier auscultatory device.17 In this study, after correction for the clearcut influence of systolic BP level, there was no significance contribution by subjects' age to the perceived quality of device performance.

**Phase III: Field Testing**  The major thrust of the study was the evaluation of factors that might confound interpretation of published validation studies. Of those, the most predictable was the likely influence of physical activity on device performance during field testing. From first principles, one might expect that the more active the participant the lesser the chance of measure-

| TABLE 3. COMPARISON OF SUBJECT CHARACTERISTICS OF THE IN-FIELD TESTING (PHASE III) AND THE STATIC VALIDATION (PHASE V) STUDY PARTICIPANTS AND THE ABPM CLINIC CONTROLS |
|----------------|----------------|----------------|------------------|------------------|------------------|
|                | SBP (mm Hg)    | DBP (mm Hg)    | Arm Circ. (cm)   | Age (years)      | Activity (units) |
| In-field testing participants (n = 21) | 119 ± 10*     | 74 ± 7*        | 26 ± 2          | 30 ± 8*          | 49 ± 15          |
| Validation study group (n = 85)       | 155 ± 42*     | 88 ± 18        | 28 ± 7          | 50 ± 18          | NA              |
| ABPM clinic controls (n = 102)        | 135 ± 15      | 83 ± 10        | 28 ± 8          | 48 ± 13          | 43 ± 13          |

* \( P < .0001 \) compared with controls, \( P < .0001 \) compared with controls.

ABPM, ambulatory blood pressure monitoring; Circ., circumference; DBP, diastolic blood pressure; NA, not applicable; SBP, systolic blood pressure.
ment rejection. A concern with ABPM is that its performance tends to be poorer under ambulatory conditions, possibly because of errors originating from movement of the arm, sliding of the cuff, contraction of muscles and, particularly in the case of auscultatory devices, environmental noise. Such was the case: both the average level of activity over the daytime and the vigor of activity at the time of measurement had significant influence on rejection rates. Accordingly, there may be a temptation in future studies to advise subjects to limit activity during this phase, or to select predominantly sedentary subjects in the interests of expediting a positive protocol outcome. With increasing recognition of the importance of objectively quantifying physical activity in the interpretation of ABPM data, it may be timely to require documentation of activity within Phase III of the protocol, and to demand evidence that levels representative of the normal population were achieved. Accordingly, we believe that electronic activity monitoring should be incorporated in the validation protocol for ABPM units. The availability of several models of wrist worn devices, with simple computerized analytical programs, make this a reasonable and achievable refinement.

In contrast to the static evaluation phase, there were significant differences between the participants in Phase III and clinic patients. In essence, the study subjects appear to have been normal volunteers. While they were representative in activity levels, the biggest determinant of performance in this phase, they exhibited significantly lower BP. If, as we have shown, systolic BP levels influence the accuracy of diastolic BP estimation and if a monitor's program algorithms include criteria for rejecting readings based on the extent of BP difference from preceding readings, then BP levels may be a determinant of outcome in field testing also. Statistical analysis of our data did not detect BP influence, however, the selected participants exhibited quite low BP compared with patient controls and compared with Phase V subjects. Although the revised protocol now requires some hypertensives in field testing, it would be wise to specify that participants in this phase should be broadly representative of clinical practice.

In conclusion, ABP monitor validation protocols should incorporate the use of electronic activity monitors to provide objective measurement of physical activity during field testing, demonstrating normal levels of activity. Protocols should require that the subjects involved in field testing of the units should be recruited from the physically active hypertensive population, whereas accuracy assessment should focus on levels of blood pressure that are relevant in clinical practice.

REFERENCES