A Comprehensive Evaluation of Rhythm Monitoring Strategies for the Detection of Atrial Fibrillation Recurrence Insights From 647 Continuously Monitored Patients and Implications for Monitoring After Therapeutic Interventions

Efstratios I. Charitos, MD; Ulrich Stierle, MD; Paul D. Ziegler, MS; Malte Baldegig, MD; Derek R. Robinson, MA, MSc, DPhil, CStat; Hans-Hinrich Sievers, MD; Thorsten Hanke, MD

Background—Intermittent rhythm monitoring (IRM) to detect atrial fibrillation (AF) recurrence is employed to evaluate the success of therapeutic interventions. In a large population of patients with continuous monitoring (CM), we investigated the sensitivity of various frequencies and durations of IRM strategies on the detection of AF recurrence, the dynamics behind AF recurrence detection, and we describe measures to evaluate temporal AF recurrence.

Methods and Results—Rhythm histories of 647 patients (mean AF burden, 0.12±0.22; median, 0.014; 687 patient-years) with implantable CM devices were reconstructed and analyzed. With the use of computationally intensive simulation, the sensitivity of IRM of various frequencies and durations on the identification of AF recurrence was evaluated. Prolonged-duration IRM was superior to shorter IRM (P<0.0001). However, even with aggressive IRM strategies, AF recurrence was not detected in a great proportion of patients. The temporal AF burden aggregation (AF density) was directly related to IRM sensitivity (P<0.0001). Even at similar AF burdens, patients with high-density AF required higher-frequency or prolonged-duration IRM to achieve the same sensitivity as in low-density AF (P<0.0001). Patients with high-density, low-burden AF benefit the most from CM for detection of AF recurrence.

Conclusions—IRM follow-up is significantly inferior to CM. IRM strategies will not identify AF recurrence in a great proportion of patients at risk. Temporal AF characteristics play a significant role in AF recurrence detection with the use of IRM. For the scientific, evidence-based evaluation of AF treatments, CM should be strongly recommended. Prospective studies are required to evaluate whether CM to guide clinical management can also improve patient outcomes.

Clinical Trial Registration—URL: http://www.clinicaltrials.gov. Unique identifier: NCT00806689.

(Circulation. 2012;126:806-814.)

Key Words: atrial fibrillation ■ atrial fibrillation arrhythmia ■ rhythm monitoring ■ rhythm recorders

Detection of atrial fibrillation (AF) recurrence after therapeutic interventions has until now been based primarily on the results of intermittent (“snapshot”) rhythm monitoring (IRM). This approach has significant limitations in terms of sensitivity because the monitoring period of these examinations is limited; however, the results of these examinations are used to draw inferences on the success of ablation procedures or pharmacological strategies. It has been shown that reliance on symptoms or office ECGs overestimates the success rate of the ablation procedures and will misclassify patients who have recurrence of paroxysmal AF as being in sinus rhythm.1–4 The current consensus on AF monitoring recommendations at least two 24-hour Holter monitor (HM) examinations annually for the detection of AF recurrence after ablation procedures,5,6 which, however, has also been shown to underdetect AF recurrence and thus overestimate procedural success.1,8 Reliable and accurate detection of AF recurrence is thus of special importance for the evaluation of pharmacological or ablation therapies as well as when decisions on changes in anticoagulation or antiarrhythmic therapy are to be made.

Editorial see p 791
Clinical Perspective on p 814

With the introduction of implantable, leadless rhythm recorders, continuous monitoring (CM) has been proposed
recently for more accurate detection of AF recurrence after ablation procedures or novel pharmacological strategies. As expected, the sensitivity of these diagnostic modalities is close to 100% and, as such, is much superior to that of the currently employed 24-hour HM strategies. However, the implantation of a CM device in all patients may seem unrealistic. This led to the belief that intensifying noninvasive IRM either with greater frequency or with longer duration (7- or 30-day monitoring) or both may lead to better and more reliable detection of AF recurrence.

Recently, data from CM patients have been used to correlate the recurrence of AF with the risk for thromboembolic complications. The advance of this technology has not only provided better insight into the rhythm evaluation of these patients but has also initiated a change of mentality regarding AF from a qualitative ("yes/no") to a quantitative (amount of AF) approach. The AF burden, defined as the proportion of the total monitored time a patient is in AF, has been utilized in evaluating the risk for thromboembolism in patients with AF. However, even these measures present a rather static approach in the evaluation of the dynamic incidence of AF recurrence, partially ignoring the temporal dispersion or temporal aggregation of the AF episodes and AF burden.

The aim of the present study was 2-fold: (1) to identify the sensitivity of IRM strategies of various frequencies and durations on the accurate detection of AF recurrence with the use of data collected from patients with heart rhythm CM devices and (2) to investigate the dynamics behind AF recurrence detection and propose measures to evaluate the temporal pattern of AF recurrence.

Methods

Data acquired from 647 patients monitored with a CM device (Reveal XT 9529, n = 73; AT500 pacemaker, n = 574; Medtronic, Inc, Minneapolis, MN) were analyzed. Demographics and patient characteristics are summarized in Table 1. All patients provided informed consent for the data collection and use, and the study was approved by the local ethics committee.

With the use of data from the CM device, all AF episodes were examined. Isolated AF episodes of <5 minutes duration that could present artifacts were disregarded, and patients with no documented AF episode were regarded as AF free (n = 174). The complete rhythm history of every patient was reconstructed (Figure 1). Thereafter, we calculated the probability of successful identification of AF recurrence by IRM of various durations (eg, 24 hours and 7, 14, and 30 days) and frequencies for every patient. Computationally intensive simulation was employed to simulate in every patient all possible IRM strategies of various durations and frequencies to draw inferences on the sensitivity of the monitoring strategy and burden characteristics of the patients being monitored. Sensitivity was defined as the proportion of patients correctly identified as having AF recurrence with the simulated IRM strategy to the true number of patients having AF recurrence (identified from CM). AF burden was defined as the proportion of the time in AF in relation to the total monitored time. The simulation trials were performed as follows: After reconstructing the rhythm history of every patient, monitored for a total of g days, we defined the sample space \( \Omega_g = \{1, 2, \ldots, g-k+1\} \) to be the set of possible days that a k-day intermittent monitoring could be started (for \( k = 1, 2, 7, 14, \) and 30). A k-day monitoring starting on day \( i \in \Omega_g \) therefore included the following associated monitored days: \( i, i+1, \ldots, i+k-1 \). To simulate \( n \) independent k-day monitorings of patient \( j \), \( n \) elements were selected at random from \( \Omega_g \), except that elements were rejected if their monitored days intersected with the monitored days of previously selected elements. AF was deemed to have been successfully identified if it was observed in at least 1 of the \( n \) sets of monitored days. This was performed for all patients of the study population, for monitoring durations of \( k = 1, 2, 7, 14, \) and 30 days, and for strategies

### Table 1. Demographics of the Patient Population

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Total</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>376</td>
<td>58.1</td>
</tr>
<tr>
<td>Age, mean±SD, y</td>
<td>68.9±12.3</td>
<td></td>
</tr>
<tr>
<td>Follow-up, mean±SD, range, y</td>
<td>1.1±0.4, 0.1–3.7</td>
<td></td>
</tr>
<tr>
<td>History of atrial arrhythmia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atrial tachycardia</td>
<td>114</td>
<td>17.6</td>
</tr>
<tr>
<td>Atrial flutter</td>
<td>176</td>
<td>27.2</td>
</tr>
<tr>
<td>Paroxysmal AF</td>
<td>475</td>
<td>73.4</td>
</tr>
<tr>
<td>Persistent AF</td>
<td>32</td>
<td>4.9</td>
</tr>
<tr>
<td>Long-lasting persistent AF</td>
<td>35</td>
<td>5.4</td>
</tr>
<tr>
<td>History of cardioversion</td>
<td>18</td>
<td>2.8</td>
</tr>
<tr>
<td>Cardiovascular history</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ischemic heart disease</td>
<td>99</td>
<td>15.3</td>
</tr>
<tr>
<td>Coronary artery disease</td>
<td>220</td>
<td>34.0</td>
</tr>
<tr>
<td>Cardiomyopathy</td>
<td>64</td>
<td>9.9</td>
</tr>
<tr>
<td>Hypertension</td>
<td>405</td>
<td>62.6</td>
</tr>
<tr>
<td>History of ablation for AF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cox/Maze III procedure</td>
<td>17</td>
<td>2.6</td>
</tr>
<tr>
<td>Left sided only</td>
<td>53</td>
<td>8.2</td>
</tr>
<tr>
<td>AV node ablation</td>
<td>28</td>
<td>4.3</td>
</tr>
<tr>
<td>Other</td>
<td>71</td>
<td>11.0</td>
</tr>
<tr>
<td>History of cardiac surgery</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CABG</td>
<td>113</td>
<td>17.5</td>
</tr>
<tr>
<td>MVR</td>
<td>45</td>
<td>7.0</td>
</tr>
<tr>
<td>AVR</td>
<td>41</td>
<td>6.3</td>
</tr>
<tr>
<td>TVR</td>
<td>7</td>
<td>1.1</td>
</tr>
<tr>
<td>Ascending aorta replacement</td>
<td>9</td>
<td>1.4</td>
</tr>
<tr>
<td>PVR</td>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>NYHA class</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>331</td>
<td>51.2</td>
</tr>
<tr>
<td>II</td>
<td>234</td>
<td>36.2</td>
</tr>
<tr>
<td>III</td>
<td>67</td>
<td>10.4</td>
</tr>
<tr>
<td>IV</td>
<td>3</td>
<td>0.5</td>
</tr>
<tr>
<td>Pacing indication</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AV block</td>
<td>85</td>
<td>13.1</td>
</tr>
<tr>
<td>Sinus node dysfunction</td>
<td>397</td>
<td>61.4</td>
</tr>
<tr>
<td>Other</td>
<td>41</td>
<td>6.3</td>
</tr>
<tr>
<td>Arrhythmia-related medication</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class I</td>
<td>89</td>
<td>13.8</td>
</tr>
<tr>
<td>Class III</td>
<td>251</td>
<td>38.8</td>
</tr>
<tr>
<td>β-blocker</td>
<td>212</td>
<td>32.8</td>
</tr>
<tr>
<td>Calcium channel blocker</td>
<td>56</td>
<td>8.7</td>
</tr>
<tr>
<td>Digoxin</td>
<td>144</td>
<td>22.3</td>
</tr>
</tbody>
</table>

AF indicates atrial fibrillation; AV, atrioventricular; CABG, coronary artery bypass grafting; MVR, mitral valve replacement/repair; AVR, aortic valve replacement/repair; TVR, tricuspid valve replacement/repair; PVR, pulmonary valve replacement; and NYHA, New York Heart Association.
The numbers 1 and 2 denote the time points of 2 random 24-hour Holter monitoring (HM) would provide. For patient B, increases the sensitivity by only 3% from that which random 24-hour Holter monitoring (HM) would provide. For patient B, the respective probabilities are Pr(24h|B)=0.23 and Pr(30d|B)=0.26. In this case, a random 30-day monitoring increases the sensitivity by only 3% from that which random 24-hour Holter monitoring (HM) would provide. For patient B, the probability of identification of AF recurrence for the development of a proportion $p$ of the patient’s total observed burden $b$ as $F(p,b)$. The absolute cumulative deviation of the patient’s actual burden development from the hypothetical uniform burden development is $1 - b/2$. AF density is numerically evaluated as follows:

$$\text{AF density} = \frac{\int_0^1 F(p,b) - p \, dp}{1 - b}$$

The detection probability gain of a CM versus an IRM strategy of $k$ duration [with a probability of AF detection in patient $i$ of $Pr(k|i)$] was evaluated as $1 - Pr(k|i)$ and depicts the increase in probability of AF detection that CM offers in comparison to that of a random IRM of $k$ duration ($k$: 24 hours, 7 days, 14 days, 30 days). Linear regression was used to determine whether AF burden and AF density were independently associated with the probability gain. To restore normality, the log of the probability gain was used as dependent variable.

The Mann-Whitney test was used for comparing AF burdens among patients with high and low density. All statistical analyses and procedures were performed with R version 2.14.1 (R Development Core Team; 2011; http://www.R-project.org/).

Results

Sensitivity of IRM

Quantitative AF characteristics are displayed in Figure 3. Figure 4 displays the results for the sensitivity of random 24-hour and 7-, 14-, and 30-day recordings at various durations to detect AF recurrence for the entire patient population. Prolonged IRM was significantly superior to shorter IRM ($P<0.0001$ for all comparisons). However, a four 24-hour HM strategy had a sensitivity of only 52%, thus failing to identify AF recurrence in almost half of the patients with proven AF recurrence. Even with a theoretical strategy of three 30-day HM tests (for a total monitoring duration of 90 days per year), the sensitivity did not exceed 82%, and thus a nonnegligible proportion of patients with AF would be misclassified (Figure 4). However, the increased sensitivity offered by prolonged IRM durations comes at a cost of disproportionate increase in the required monitored time to achieve that level of sensitivity. For a sensitivity of 0.65, 30 days of monitoring would have been required with a 30-day
HM strategy (one 30-day HM test), whereas for the same sensitivity, 7 monitored days would have been required with a 24-hour HM strategy (seven 24-hour HM tests) (Figure 4, solid horizontal black line).

Inferences Drawn From a Series of Negative IRM
Using computationally intensive simulation, we sought to evaluate the inferences that can be drawn from a series of negative IRM on the patients’ AF burden. Although IRM strategies cannot determine the precise AF burden, reliable inferences can be obtained (for example, a patient with 4 negative random 30-day HM tests is highly unlikely to have a burden of >80%). These probabilities were derived from our patient population, taking into consideration not only the amount of AF in each patient but also the temporal characteristics of AF, and are displayed in Figure 5. A series of 4 negative random 30-day HM tests suggests with 90% confidence that this patient’s AF burden is 17%. Similarly, 8 negative 24-hour HM tests would be required to achieve the same level of confidence regarding the patient’s AF burden. IRM of shorter duration requires a higher number of examinations to achieve the same level of confidence as IRM of longer durations (Figure 5).

Influence of AF Burden and AF Density on Effectiveness on IRM
Using the reconstructed rhythm histories of our patients, we attempted to evaluate the quantitative as well as temporal AF characteristics that drive the sensitivity differences between IRM of various durations and frequencies. Burden density was directly related to the sensitivity of the different monitoring durations (Figure 6). Even with similar AF burdens (burden of low-density group: mean, 0.12±0.11; median, 0.07; quartile 1, 0.01; quartile 3, 0.18; burden of high-density group: mean, 0.13±0.14, median, 0.06; quartile 1, 0.01; quartile 3, 0.19; P=0.36), patients with low-density AF (AF density <0.5) achieve higher sensitivities for AF recurrence detection compared with patients with high-density AF (AF density >0.5; P<0.0001 for all comparisons of all monitoring duration sensitivities [24-hour, 7-day, 14-day, 30-day HM] between high- and low-density groups). In patients with high-density AF, AF recurrence is much more difficult to

Figure 2. Two patients with the same atrial fibrillation (AF) burden but different burden aggregation (patient C, high-density AF; patient D, low-density AF). With the rhythm history reconstructed, the course of the minimum monitored time required for each burden proportion is plotted against the proportion of the total burden required (dotted lines, patients C and D, bottom). Patient C developed 50% of his total burden in 11% of the monitored time (black dot, patient C, bottom). Patient D, in contrast, required 40% of the observation time to develop 50% of his burden (black dot, patient D, bottom) as each day contributes less to the total burden because the AF burden is spread over more days. The black diagonal line (patients C and D, bottom) represents a hypothetical uniform AF burden development. The area between the actual (blue or red dotted line) and the uniform hypothetical (solid black diagonal) AF burden development is evaluated as a measure of the temporal aggregation of the AF burden (AF burden density).

Figure 3. Histogram of the patients’ atrial fibrillation (AF) burdens. A total of 174 patients with no documented AF episode during the monitored time were excluded.
by guest on April 20, 2017 http://circ.ahajournals.org/ Downloaded from

Previously,10,17–19 but to evaluate the characteristics and dynamics of our study was not to evaluate the AF detection process of the devices, which has been extensively validated previously. Although 2 different monitoring devices were employed, AF detection in both devices was similar. The aim of our study was not to evaluate the AF detection process of the devices, which has been extensively validated previously.10,17–19 but to evaluate the characteristics and dynamics of AF recurrence detection after AF episode registration.

**Discussion**

This study provides insights into the sensitivity of various monitoring strategies for the detection of AF recurrence with the use of data from a large patient population monitored with CM devices. Although 2 different monitoring devices were employed, AF detection in both devices was similar. The aim of our study was not to evaluate the AF detection process of the devices, which has been extensively validated previously.10,17–19 but to evaluate the characteristics and dynamics of AF recurrence detection after AF episode registration.

**Figure 4.** The sensitivity of intermittent rhythm monitoring (IRM) to detect atrial fibrillation (AF) recurrence as a function of monitoring frequency (number of random monitorings) and monitoring duration (24 hours and 7, 14, and 30 days) in patients with proven AF recurrence. Only patients with documented AF recurrence were included. Prolonged IRM was significantly superior to shorter IRM (P<0.0001 for all comparisons). However, a four 24-hour Holter monitoring (HM) strategy had a sensitivity of only 52%, thus failing to identify AF recurrence in almost half of the patients with proven AF recurrence. The dotted horizontal line represents the sensitivity of 0.5 (fair coin). To achieve a sensitivity of 0.65 (solid horizontal line), 30 days of monitoring would have been required with a 30-day HM strategy (one 30-day HM test). To achieve the same sensitivity, 7 monitored days would have been required with a 24-hour HM strategy (seven 24-hour HM tests). The total number of monitored days required at the maximum frequency of each monitoring strategy is denoted.

**Figure 5.** Inferences on atrial fibrillation (AF) burden drawn by hypothetical negative intermittent rhythm monitoring (IRM) examinations of the overall study population. For the study population, a series of 2 negative random 24-hour Holter monitor (HM) tests imply with a confidence of 90% that the patient’s burden is ≤32%. Similarly, a series of 4 negative random 30-day HM tests imply with a confidence of 90% a burden of ≥17%. Eight negative series of 24-hour HM tests would have been required to obtain the same level of AF burden confidence.

Although this patient population presents an inhomogeneous collection of patients being monitored continuously with implantable devices, this gives us the opportunity to observe the characteristics of a very wide range of AF burden spectrum. The lower end of the AF burden spectrum is predominantly occupied by patients implanted with a pacemaker device for monitoring after ablation procedures (mean AF burden, 0.05±0.16; median, 0.002; quartile 1, 0.001; quartile 3, 0.01) or patients with paroxysmal, asymptomatic AF, whereas the higher end of the spectrum is predominantly occupied by patients implanted with a pacemaker device for symptomatic bradycardia and long-lasting persistent AF. This wide range of AF burden observed allowed us to evaluate the sensitivity of the IRM strategies and evaluate the AF burden dynamics over a wide range of AF burdens.

**Judgment of Therapeutic Success Based on IRM**

Our group and other groups have previously evaluated sensitivity differences between IRM and CM by comparing the information obtained from CM devices with actual IRM performed in the same patients.14,8,11 However, these methods have limitations because the time point of the IRM plays a major role in the evaluation of these methods. As depicted in Figure 1, in patient A, a 24-hour HM test performed at time point 1 will be negative and patient A will be regarded as “AF free,” whereas the same examination performed some days earlier at time point 2 would have identified AF recurrence. The results of such studies comparing continuous versus actual IRM are severely influenced by the choice of time points when the IRM is performed. Because in most of these studies only a limited number of IRM tests were compared.
against CM, chance has an immeasurable effect on the results of such studies, and any inferences drawn from these results will be problematic. In the present study, the reconstruction of the rhythm history of every patient allows us to evaluate the true probability of AF recurrence detection of any random IRM of any duration in every patient, and computationally intensive simulation can draw inferences on the sensitivities and success rates of IRM of any duration and frequency.

Figure 4 presents the results of this approach in our patient collective, depicting the sensitivities of the 4 most common IRM durations at various frequencies in patients with high (>0.5) and low (<0.5) atrial fibrillation (AF) burden density. Both high- and low-density AF groups had similar AF burdens (burden of low-density group: mean, 0.12 ± 0.11; median, 0.07; quartile 1, 0.01; quartile 3, 0.18; burden of high-density group: mean, 0.13 ± 0.14; median, 0.06; quartile 1, 0.01; quartile 3, 0.19; P = 0.36). Low-density AF patients achieve higher AF recurrence detection sensitivity with less frequent monitoring than patients with high-density AF. In patients with high-density AF (blocks of AF), a significantly (P < 0.0001) higher monitoring frequency is required to detect recurrence of AF. Only patients with documented AF recurrence were included. HM indicates Holter monitoring.

Table 2. Linear Regression Analysis of Effect of Burden and Burden Density on the Probability Gain of AF Recurrence Detection of 30-Day and 24-Hour HM vs Continuous Monitoring

<table>
<thead>
<tr>
<th>Factor</th>
<th>Coefficient</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain in AF recurrence detection:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>continuous vs 24-h HM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF burden</td>
<td>-5.76</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>AF burden density</td>
<td>2.61</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Gain in AF recurrence detection:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>continuous vs 30-d HM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF burden</td>
<td>-4.08</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>AF burden density</td>
<td>4.84</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

AF indicates atrial fibrillation; HM, Holter monitoring. To restore normality, the log of the probability gain was used as dependent variable.

Figure 6. Sensitivity of various intermittent rhythm monitoring durations at various frequencies in patients with high (>0.5) and low (<0.5) atrial fibrillation (AF) burden density. Both high- and low-density AF groups had similar AF burdens (burden of low-density group: mean, 0.12 ± 0.11; median, 0.07; quartile 1, 0.01; quartile 3, 0.18; burden of high-density group: mean, 0.13 ± 0.14; median, 0.06; quartile 1, 0.01; quartile 3, 0.19; P = 0.36). Low-density AF patients achieve higher AF recurrence detection sensitivity with less frequent monitoring than patients with high-density AF. In patients with high-density AF (blocks of AF), a significantly (P < 0.0001) higher monitoring frequency is required to detect recurrence of AF. Only patients with documented AF recurrence were included. HM indicates Holter monitoring.

against CM, chance has an immeasurable effect on the results of such studies, and any inferences drawn from these results will be problematic. In the present study, the reconstruction of the rhythm history of every patient allows us to evaluate the true probability of AF recurrence detection of any random IRM of any duration in every patient, and computationally intensive simulation can draw inferences on the sensitivities and success rates of IRM of any duration and frequency.

Figure 4 presents the results of this approach in our patient collective, depicting the sensitivities of the 4 most common IRM durations (24-hour and 7-, 14-, and 30-day HM) at various monitoring frequencies. In our patient population, a monitoring strategy with four 24-hour HM tests would have a sensitivity of only 52%, thus failing to identify recurrence of AF in almost half of these patients. This low sensitivity becomes even more striking if one considers that the flipping of a fair coin would have provided the same sensitivity for the detection of AF recurrence in these patients. The direct implication of these results is that the evaluation of pharmacological or interventional therapies when a 24-hour HM is employed as a monitoring strategy should be interpreted cautiously because a great proportion of patients will be misclassified, AF recurrence will be underdiagnosed, and therapeutic success will be overestimated. Similarly, a single, random 30-day HM in our patient population will have a sensitivity of 63%, and thus the evaluation of the success of an ablation procedure with the use of a single, postinterven-
intermittent monitoring altogether, a great part or even the majority of patients will be misclassified, and therefore the scientific evaluation of therapeutic interventions for AF becomes problematic.

**Semiquantitative Estimation of AF Burden Based on Negative IRM**

Although IRM may fail to detect AF recurrence in a great proportion of patients and cannot provide the level of information a CM strategy does, results of serial IRM may guide patient management in a semiquantitative way when CM is not possible because of either limited availability, patient compliance, or cost considerations. Such an evaluation is depicted in Figure 5. A series of negative IRM, although may underestimate qualitative AF detection per se, can however guide estimation of the potential AF burden level in patients prone to AF recurrence such as patients after ablation procedures, or patients on rhythm control strategies. In our large population of CM patients, 4 random negative 24-hour HM tests indicate, with a confidence of 90%, an AF burden of up to 23%. A higher number of serial negative IRM tests or negative IRM tests of greater duration can pinpoint the potential AF burden to lower levels. However, it is important to note that even average burdens that low have been shown to be a risk factor for thromboembolic complications.13–16,21

**AF Characteristics Influencing the Success of IRM**

A major variable driving the sensitivity in detecting AF recurrence is not only the AF burden but the dispersion of it over time. AF density, as a measure of this dispersion, proved to be a major factor that influences the sensitivity of IRM at various frequencies and durations. Although patients A and B of Figure 1 have very similar amounts of AF burden over the same observation time, the probability of AF recurrence detection is vastly different because of the different temporal AF burden dispersion. In patient B (Figure 1), because of the even dispersion of AF burden, prolonged-duration IRM (30-day HM) is not superior to shorter-duration IRM (24-hour HM); the probability of AF recurrence detection of both IRM tests is 1, and the sensitivity of both strategies is 100%. In patient A (Figure 1), however, because of the high AF density, a random 30-day HM test offers only a negligible 3% increase in probability of AF recurrence detection compared with a random 24-hour HM test. In the whole patient cohort (Figure 6), the AF density appears to have a major and statistically significant influence on the sensitivity of AF detection with the use of IRM strategies. In patients with high-density AF, the probabilities of AF recurrence detection and the sensitivities of IRM strategies are significantly inferior to those of patients with same level of AF burden but low AF density (Figure 6).

**Implications for Clinical Outcomes, Patient Management, and Current Knowledge**

Although the level of AF burden has been shown to be a significant risk factor for thromboembolic events,13–16 to the best of our knowledge, no study has taken into consideration the temporal characteristics of AF burden development. It is conceivable that the temporal distribution of AF burden may also play a role in the general thromboembolic risk that patients with AF face. Patients A and B (Figure 1), although spending the same time in AF, may have a different thromboembolic risk because of the different temporal distribution of their AF burden. Larger studies are required to evaluate this hypothesis, and prospective studies are required to evaluate the clinical impact of the combined quantitative and temporal AF characteristics on patient outcome.

The data presented may have some clinical implications. The results of our study depict the major limitations of IRM not only for management of patients but also for evaluation of the results of pharmacological as well as interventional therapies for AF. From a clinical perspective, accurate patient management can only be obtained with confidence with CM. From a scientific perspective, our evaluation until now of the success of pharmacological and interventional procedures based on IRM seems problematic in light of the results presented in this study. The accurate evaluation of therapies for AF in the era of evidence-based medicine mandates the use of CM.

It is conceivable that the widespread use of CM or the results from large studies of CM patients may require us to reevaluate our definitions of success or failure of therapeutic intervention for AF. It may very well be that the current definition of success of a therapeutic intervention only in terms of complete absence of AF recurrence might be inadequate and that a quantitative approach to AF recurrence may in the future require us to critically revisit our current knowledge and reevaluate our therapeutic interventions and
strategies, our results, and our patient management decision-making process. Whether the use of CM to guide clinical management can also improve patient outcomes remains uncertain at this point and must be investigated in large randomized controlled prospective studies.

Limitations
This study was underpowered to investigate the clinical impact of temporal AF characteristics on patient outcomes. We believe that this should be the scope of large prospective studies or registries of CM patients over sufficient follow-up time. AF episodes of patients primarily implanted with a CM device because of atrial tachycardia or atrial flutter have been included in the analysis. In our experience, although patients are typically classified in 1 distinct rhythm type (atrial flutter, AF, or atrial tachycardia), CM reveals that these patients actually tend to experience multiple types of atrial rhythms. Although 2 different devices were employed in the population of this study, previous studies have extensively investigated the accuracy of AF detection of these devices (sensitivity, ≥98%; specificity, ≥91%). Additionally, we tried to manually inspect all available episodes and, in cases of uncertainty or dispute, we discarded isolated, short (<5 minutes) episodes that could represent artifacts.

Conclusion
Patient follow-up with either short- or long-duration IRM is significantly inferior to CM for the detection of AF recurrence. IRM monitoring strategies will not identify AF recurrence in a great proportion of patients at risk. With short-duration, low-frequency IRM strategies, chance has an immeasurable effect on the evaluation of AF recurrence. Thus, results of pharmacological or invasive interventions evaluated with IRM should be considered with caution. Temporal AF characteristics play a major role in AF recurrence detection with the use of IRM strategies and may influence clinical parameters and outcomes in AF patients. For the scientific, evidence-based evaluation of AF treatments, CM should be strongly recommended.

Acknowledgments
We are thankful to Katrin Meyer and Martina Schröder for secretarial support and to Jana Paise, Anja Paap, and Tobias Frin for data documentation support.

Disclosures
Dr Charitos, Stierle, Baldewig, Robinson, and Sievers have no conflict of interest to disclose. Paul D. Ziegler is a full-time employee of Medtronic (USD). Dr Hanke discloses modest lecture honoraria (10 000 USD) and stockholder of Medtronic (USD). Dr Hanke discloses modest lecture honoraria (10 000 USD) and stockholder of Medtronic (USD). Dr Hanke discloses modest lecture honoraria (10 000 USD) and stockholder of Medtronic (USD).

References


---

**CLINICAL PERSPECTIVE**

Patient follow-up with either short- or long-duration intermittent rhythm monitoring is significantly inferior to that with continuous rhythm monitoring for the detection of atrial fibrillation (AF) recurrence. Intermittent rhythm monitoring strategies fail to identify AF recurrence in a great proportion of patients at risk. Even with feasible, aggressive intermittent monitoring strategies (such as quarterly 24-hour Holter monitoring), AF recurrence will not be identified in a significant proportion of patients. Because of the low sensitivity of intermittent rhythm monitoring, the results of pharmacological or invasive interventions for AF when evaluated with intermittent monitoring should be considered with caution because chance has an immeasurable effect on the detection of AF recurrence, and thus a great proportion of patients will be misclassified. AF recurrence will be underdiagnosed, and therapeutic success will be overestimated. Novel quantitative and temporal AF characteristics (AF burden, AF density) play a major role in AF recurrence detection with the use of intermittent monitoring strategies and may influence clinical parameters and outcomes in AF patients. For accurate patient management, as well as for the scientific, evidence-based evaluation of AF treatments, continuous rhythm monitoring should be strongly recommended.

---

Go to http://cme.ahajournals.org to take the CME quiz for this article.
A Comprehensive Evaluation of Rhythm Monitoring Strategies for the Detection of Atrial Fibrillation Recurrence: Insights From 647 Continuously Monitored Patients and Implications for Monitoring After Therapeutic Interventions

Efstratios I. Charitos, Ulrich Stierle, Paul D. Ziegler, Malte Baldewig, Derek R. Robinson, Hans-Hinrich Sievers and Thorsten Hanke

_Circulation_. 2012;126:806-814; originally published online July 23, 2012; doi: 10.1161/CIRCULATIONAHA.112.098079

_Circulation_ is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2012 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/126/7/806

An erratum has been published regarding this article. Please see the attached page for:
/content/126/13/e216.full.pdf

Data Supplement (unedited) at:
http://circ.ahajournals.org/content/suppl/2013/10/14/CIRCULATIONAHA.112.098079.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in _Circulation_ can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to _Circulation_ is online at:
http://circ.ahajournals.org/subscriptions/
In the article by Charitos et al., “A Comprehensive Evaluation of Rhythm Monitoring Strategies for the Detection of Atrial Fibrillation Recurrence: Insights From 647 Continuously Monitored Patients and Implications for Monitoring After Therapeutic Interventions” which was published in the August 14, 2012 issue of the journal (Circulation. 2012;126:806–814), an error occurred in the legend to Figure 1. The legend should read:

“Reconstruction of the rhythm history of patients A and B. Both patients have similar atrial fibrillation (AF) burden (AF burden for patient A=0.22; AF burden for patient B=0.21). The vastly different temporal aggregation of the AF episodes and AF burden in these patients is responsible for the different sensitivities of intermittent AF monitoring modalities in these patients.

For patient A, the probability of identification of AF recurrence in a 24-hour and 30-day monitor is Pr(24h|A)=0.23 and Pr(30d|A)=0.26. In this case, a random 30-day monitoring increases the sensitivity by only 3% from that which random 24-hour Holter monitoring (HM) would provide.

For patient B, the respective probabilities are Pr(24h|B)=1 and Pr(30d|B)=1.

The numbers 1 and 2 denote the time points of 2 random 24-hour HM tests, whereas the numbers 3 and 4 denote the time points of 2 random 30-day HM tests.”

The current online version of the article has been corrected. The authors regret the error.
심방세동 재발을 감지하기 위해서는 연속적인 모니터링이 필요하다

오 세일 교수 서울대학교병원 순환기내과

Summary

배경
심방세동 재발을 감지하기 위한 간헐적인 리듬 모니터링(intermittent rhythm monitoring, IRM)은 시술 후 성공 여부를 평가하기 위한 수단으로 사용되고 있다. 연속적인 모니터링(continuous monitoring, CM)을 받은 대규모 환자군 자료를 이용하여 심방세동 재발을 감지하기 위한 IRM 전략의 비도 및 기간에 따른 민감도와 심방세동 재발에 깔린 역동학을 조사하였고, 시간적인 심방세동 재발 평가를 위한 수단을 기술하였다.

방법 및 결과
이식형 CM 장치를 가진 647명의 환자들(평균 심방세동 부담, 0.12±0.22; 중앙값, 0.014; 687명 환자-년)의 리듬 내역을 분석하였다. 전산 집중 시뮬레이션을 이용하여 심방세동 재발 감지를 위한 IRM의 비도와 기간에 따른 민감도를 평가하였다. 기간이 긴 IRM이 짧은 경우보다 우월하였다 (P<0.0001). 그러나 공격적인 IRM 전략으로도 많은 수의 환자에서는 심방세동의 재발을 감지하지 못했다. 시간적인 심방세동 부담의 밀집(심방세동 밀도)은 IRM의 민감도와 직접적으로 연관되어 있었다 (P<0.0001).심지어 유사한 심방세동 부담에서도 고밀도의 심방세동을 가진 환자의 경우, 저밀도 심방세동을 가진 환자와 같은 민감도를 갖기 위해서는 더 자주 또는 더 오랜 기간의 IRM이 요구된다 (P<0.0001). CM은 고밀도이면서 부담이 낮은 심방세동 환자에서 심방세동 재발 감지에 가장 도움이 될 수 있다.

결론
IRM을 이용한 추적관찰은 CM보다 유의하게 열등하며,IRM 전략은 많은 수의 환자에서 심방세동 재발을 감지하지 못할 것이다. 시간적인 심방세동의 특징은 IRM 사용시 심방세동 감지에 중요한 역할을 한다. 과학적이고 근거 중심적인 심방세동 치료의 평가를 위해서는 CM이 강력하게 권장되는데, CM이 환자의 성적(outcome)을 항상시킬 수 있는지 평가하기 위해서는 전향적인 연구가 필요하다.
심방세동의 발생은 과거부터 약물치료 및 동리듬 전환술의 성적을 평가하는 데 중요한 잣대였다. 특히, 재발 감지의 중요성은 전극도자절제술이 활발하게 적용되기 시작하면서 부각되었다. 이는 시술의 성적을 평가함에 있어서 가장 중요한 종료점 중의 하나이기 때문이다. 따라서 시술 성적이 너무 좋은 임상연구의 경우 추적관찰의 방법에 문제 제기를 받는 경우가 많았다. 집중적인 모니터링을 하는 연구일수록 심방세동 재발을 보다 잘 감지하기 때문에 성적은 낮게 나오기 마련이다. 이에 미국 및 유럽의 부정맥 학회가 공동으로 전극도자절제술과 관련된 가이드라인을 공포하였으며, 이에는 시술 후 추적관찰의 방법도 포함되어 있다. 즉, 발작성 심방세동의 경우 적어도 (1) 매 방문마다 12유도 심전도; (2) 추적 종료 시점(예, 12개월)에 24시간 헴터; (3) 시술 후 첫 3개월(blanking period)부터 추적 종료 시점까지 정기적으로 그리고 증상이 있을 때마다 사건 기록기 검사를 해야 한다고 권고한다. 하지만 위의 방법이 비교적 공격적인 모니터링이긴 하지만, Charitos 등의 본 연구 결과에 의하면 여전히 미흡하다고밖에 할 수 없다. 즉, 지금까지 알고 있는 항부정맥제나 전극도자절제술의 성적은 실제보다 과장된 것이다. 그렇다고 이 연구에서처럼 모든 환자에게 이식형 사건 기록기를 이식하는 것은 쉬운 일이 아 니다. 우리나라의 경우를 보더라도 2012년 현재 건강보험을 받지 못하고 있는 장비를 단순히 약 또는 시술에 대한 반응이 좋은지 확인하기 위한 목적으로 이식하는 것에 대해서 환자들의 인식은 매우 부정적이다. 하지만 이 연구의 결과들이 모두 실제 현실과 괴리된 것은 아니다. 임상에 적용할 수 있는 부분은 심방세동의 '부담'뿐 아니라 심방세동의 (시간적)‘밀도’ 개념이다. 각 환자에서 심방세동의 부담과 밀도 데이터를 적절히 얻어낸다면 환자별로 특화된 추적관찰 방법을 적용할 수 있을 것이다.

Reference
A Comprehensive Evaluation of Rhythm Monitoring Strategies for the Detection of Atrial Fibrillation Recurrence Insights From 647 Continuously Monitored Patients and Implications for Monitoring After Therapeutic Interventions

Efstratios I. Charitos, MD; Ulrich Stierle, MD; Paul D. Ziegler, MS; Malte Baldewig, MD; Derek R. Robinson, MA, MSc, DPhil, CStat; Hans-Hinrich Sievers, MD; Thorsten Hanke, MD

Background—Intermittent rhythm monitoring (IRM) to detect atrial fibrillation (AF) recurrence is employed to evaluate the success of therapeutic interventions. In a large population of patients with continuous monitoring (CM), we investigated the sensitivity of various frequencies and durations of IRM strategies on the detection of AF recurrence, the dynamics behind AF recurrence detection, and we describe measures to evaluate temporal AF recurrence.

Methods and Results—Rhythm histories of 647 patients (mean AF burden, 0.12±0.22; median, 0.014; 687 patient-years) with implantable CM devices were reconstructed and analyzed. With the use of computationally intensive simulation, the sensitivity of IRM of various frequencies and durations on the identification of AF recurrence was evaluated. Prolonged-duration IRM was superior to shorter IRM (P<0.0001). However, even with aggressive IRM strategies, AF recurrence was not detected in a great proportion of patients. The temporal AF burden aggregation (AF density) was directly related to IRM sensitivity (P<0.0001). Even at similar AF burdens, patients with high-density AF required higher-frequency or prolonged-duration IRM to achieve the same sensitivity as in low-density AF (P<0.0001). Patients with high-density, low-burden AF benefit the most from CM for detection of AF recurrence.

Conclusions—IRM follow-up is significantly inferior to CM. IRM strategies will not identify AF recurrence in a great proportion of patients at risk. Temporal AF characteristics play a significant role in AF recurrence detection with the use of IRM. For the scientific, evidence-based evaluation of AF treatments, CM should be strongly recommended. Prospective studies are required to evaluate whether CM to guide clinical management can also improve patient outcomes.

Clinical Trial Registration—URL: http://www.clinicaltrials.gov. Unique identifier: NCT00806689.

Key Words: atrial fibrillation ■ atrial fibrillation arrhythmia ■ rhythm monitoring ■ rhythm recorders

Detection of atrial fibrillation (AF) recurrence after therapeutic interventions has until now been based primarily on the results of intermittent (“snapshot”) rhythm monitoring (IRM). This approach has significant limitations in terms of sensitivity because the monitoring period of these examinations is limited; however, the results of these examinations are used to draw inferences on the success of ablation procedures or pharmacological strategies. It has been shown that reliance on symptoms or office ECGs overestimates the success rate of the ablation procedures and will misclassify patients who have recurrence of paroxysmal AF as being in sinus rhythm.1–4 The current consensus on AF monitoring recommends at least two 24-hour Holter monitor (HM) examinations annually for the detection of AF recurrence after ablation procedures,5–7 which, however, has also been shown to underdetect AF recurrence and thus overestimate procedural success.1,8 Reliable and accurate detection of AF recurrence is thus of special importance for the evaluation of pharmacological or ablation therapies as well as when decisions on changes in anticoagulation or antiarrhythmic therapy are to be made.

Clinical Perspective on p 46

With the introduction of implantable, leadless rhythm recorders, continuous monitoring (CM) has been proposed...
recently for more accurate detection of AF recurrence after ablation procedures or novel pharmacological strategies. As expected, the sensitivity of these diagnostic modalities is close to 100% and, as such, is much superior to that of the currently employed 24-hour HM strategies. However, the implantation of a CM device in all patients may seem unrealistic. This led to the belief that intensifying noninvasive IRM either with greater frequency or with longer duration (7- or 30-day monitoring) or both may lead to better and more reliable detection of AF recurrence.

Recently, data from CM patients have been used to correlate the recurrence of AF with the risk for thromboembolic complications. The advance of this technology has not only provided better insight into the rhythm evaluation of these patients but has also initiated a change of mentality regarding AF from a qualitative ("yes/no") to a quantitative (amount of AF) approach. The AF burden, defined as the proportion of the total monitored time a patient is in AF, has been utilized in evaluating the risk for thromboembolism in patients with AF. However, even these measures present a rather static approach in the evaluation of the dynamic incidence of AF recurrence, partially ignoring the temporal dispersion or temporal aggregation of the AF episodes and AF burden.

The aim of the present study was 2-fold: (1) to identify the sensitivity of IRM strategies of various frequencies and durations on the accurate detection of AF recurrence with the use of data collected from patients with heart rhythm CM devices and (2) to investigate the dynamics behind AF recurrence detection and propose measures to evaluate the temporal pattern of AF recurrence.

**Methods**

Data acquired from 647 patients monitored with a CM device (Reveal XT 9529, n = 73; AT500 pacemaker, n = 574; Medtronic, Inc, Minneapolis, MN) were analyzed. Demographics and patient characteristics are summarized in Table 1. All patients provided informed consent for the data collection and use, and the study was approved by the local ethics committee.

With the use of data from the CM device, all AF episodes were examined. Isolated AF episodes of <5 minutes duration that could present artifacts were disregarded, and patients with no documented AF episode were regarded as AF free (n = 174). The complete rhythm history of every patient was reconstructed (Figure 1). Thereafter, we calculated the probability of successful identification of AF recurrence by IRM of various durations (eg, 24 hours and 7, 14, and 30 days) and frequencies for every patient. Computationally intensive simulation was employed to simulate in every patient all possible IRM strategies of various durations and frequencies to draw inferences on the sensitivity of the monitoring strategy and burden characteristics of the patients being monitored. Sensitivity was defined as the proportion of patients correctly identified as having AF recurrence with the simulated IRM strategy to the true number of AF recurrence episodes and AF burden.

Table 1. Demographics of the Patient Population

<table>
<thead>
<tr>
<th>Total</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>376</td>
</tr>
<tr>
<td>Age, mean±SD, y</td>
<td>68.9±12.3</td>
</tr>
<tr>
<td>Follow-up, mean±SD, range, y</td>
<td>1.1±0.4, 0.1–3.7</td>
</tr>
</tbody>
</table>

AF indicates atrial fibrillation; AV, atrioventricular; CABG, coronary artery bypass grafting; MVR, mitral valve replacement/repair; AVR, aortic valve replacement/repair; TVR, tricuspid valve replacement/repair; PVR, pulmonary valve replacement; and NYHA, New York Heart Association.

The simulation trials were performed as follows: After reconstructing the rhythm history of every patient j, monitored for a total of g days, we defined the sample space $\Omega_g = \{1, 2, \ldots, g\}$ to be the set of possible days that a k-day intermittent monitoring could be started for (k = 1, 2, 7, 14, and 30). A k-day monitoring starting on day i in $\Omega_g$ therefore included the following associated monitored days: i, i+1, ... , i+k. To simulate n independent k-day monitored intervals of patient j, n elements were selected at random from $\Omega_g$ except that elements were rejected if their monitored days intersected with the monitored days of previously selected elements. AF was deemed to have been successfully identified if it was observed in at least 1 of the n sets of monitored days. This was performed for all patients of the study population, for monitoring durations of k = 1, 2, 7, 14, and 30 days, and for strategies
of n = 1, 2, 3, ..., 12 monitorings. The simulations were performed enough times (>50 000) to allow stabilization of the inferred parameter. All simulated IRM was of the continuous recording type. Randomization tests were used to compare sensitivities of prolonged IRM with those of shorter IRM strategies and to compare patients with high and low AF densities. The P values of 2-sided tests are reported.

Using computationally intensive simulation, we sought to evaluate the differences that can be drawn from a series of negative IRM on the patients’ AF burden. Upper 90% confidence intervals for the AF burden of patients with a series of negative examinations for a given IRM strategy were found by performing Monte Carlo simulations of the strategy and evaluating the 90th percentile of AF burden among those patients with a series of negative examinations in their simulated monitorings.

With the complete rhythm history of every patient reconstructed, we evaluated measures to describe the AF recurrence and burden development over time, taking into consideration the temporal aggregation of the episodes and the temporal dispersion of the AF burden over the monitored period. For each patient, the course of the AF burden development over time throughout the monitored period was analyzed, and the minimum monitored time required for the development of each proportion of the patient’s total observed AF burden throughout the monitored period from a hypothetical uniform burden development (area between the actual [dotted line] and the uniform burden development curves [solid diagonal], Figure 2) was calculated (blue or red highlighted area, Figure 2, bottom). This measurement was then scaled relative to the maximum possible burden aggregation for that specific patient’s burden (ie, the complete burden as 1 continuous episode (“block of AF”)) to derive the burden density, an index taking values between 0 (meaning AF burden spread evenly over the observation time) and 1 (maximum possible AF burden aggregation (ie, “1 block/episode of AF’’)).

For the numeric evaluation of the patient’s AF density, we use the following definitions: For a patient with a total AF burden b, denote the minimum proportion of contiguous monitored time required for the development of a proportion p of the patient’s total observed burden b as F(p;b). The absolute cumulative deviation of the patient’s actual burden development from the hypothetical uniform burden development is evaluated as \( f(p) = F(p;b) - p \). When the burden b occurs with maximum temporal aggregation (ie, the complete burden as 1 continuous AF episode), the minimum time required for the development of 100% of the total burden is numerically equal to b, and the cumulative deviation from the hypothetical uniform burden development is \( -\frac{1}{2} \). AF density is numerically evaluated as follows:

\[
AF
density = \frac{1 - \frac{f(p)}{b}}{1 - b}
\]

The detection probability gain of a CM versus an IRM strategy of k duration (with a probability of AF detection in patient i of Pr(k|ii)) was evaluated as \( 1 - Pr(k) - Pr(k|i) \) and depicts the increase in probability of AF detection that CM offers in comparison to that of a random IRM of k duration (k: 24 hours, 7 days, 14 days, 30 days). Linear regression was used to determine whether AF burden and AF density were independently associated with the probability gain. To restore normality, the log of the probability gain was used as dependent variable.

The Mann-Whitney test was used for comparing AF burdens among patients with high and low density. All statistical analyses and procedures were performed with R version 2.14.1 (R Development Core Team; 2011; http://www.R-project.org/).

**Results**

**Sensitivity of IRM**

Quantitative AF characteristics are displayed in Figure 3. Figure 4 displays the results for the sensitivity of random 24-hour and 7-, 14-, and 30-day monitorings at various frequencies to detect AF recurrence for the entire patient population. Prolonged IRM was significantly superior to shorter IRM (<0.0001 for all comparisons). However, a four 24-hour HM strategy had a sensitivity of only 52%, thus failing to identify AF recurrence in almost half of the patients with proven AF recurrence. Even with a theoretical strategy of three 30-day HM tests (for a total monitoring duration of 90 days per year), the sensitivity did not exceed 82%, and thus a nonnegligible proportion of patients with AF would be misclassified (Figure 4). However, the increased sensitivity offered by prolonged IRM durations comes at a cost of disproportionate increase in the required monitored time to achieve that level of sensitivity. For a sensitivity of 0.65, 30 days of monitoring would have been required with a 30-day
HM strategy (one 30-day HM test), whereas for the same sensitivity, 7 monitored days would have been required with a 24-hour HM strategy (seven 24-hour HM tests) (Figure 4, solid horizontal black line).

**Inferences Drawn From a Series of Negative IRM**

Using computationally intensive simulation, we sought to evaluate the inferences that can be drawn from a series of negative IRM on the patients' AF burden. Although IRM strategies cannot determine the precise AF burden, reliable inferences can be obtained (for example, a patient with 4 negative random 30-day HM tests is highly unlikely to have a burden of $>80\%$). These probabilities were derived from our patient population, taking into consideration not only the amount of AF in each patient but also the temporal characteristics of AF, and are displayed in Figure 5. A series of 4 negative random 30-day HM tests suggests with 90% confidence that this patient's AF burden is $\leq 17\%$. Similarly, 8 negative 24-hour HM tests would be required to achieve the same level of confidence regarding the patient's AF burden.

IRM of shorter duration requires a higher number of examinations to achieve the same level of confidence as IRM of longer durations (Figure 5).

**Influence of AF Burden and AF Density on Effectiveness on IRM**

Using the reconstructed rhythm histories of our patients, we attempted to evaluate the quantitative as well as temporal AF characteristics that drive the sensitivity differences between IRM of various durations and frequencies. Burden density was directly related to the sensitivity of the different monitoring durations (Figure 6). Even with similar AF burdens (burden of low-density group: mean, 0.12±0.11; median, 0.07; quartile 1, 0.01; quartile 3, 0.18; burden of high-density group: mean, 0.13±0.14; median, 0.06; quartile 1, 0.01; quartile 3, 0.19; P=0.36), patients with low-density AF (AF density $<0.5$) achieve higher sensitivities for AF recurrence detection compared with patients with high-density AF (AF density $>0.5$; P<0.0001 for all comparisons of all monitoring duration sensitivities [24-hour, 7-day, 14-day, 30-day HM] between high- and low-density groups). In patients with high-density AF, AF recurrence is much more difficult to

![Figure 2](image1.png)

**Figure 2.** Two patients with the same atrial fibrillation (AF) burden but different burden aggregation (patient C, high-density AF; patient D, low-density AF). With the rhythm history reconstructed, the course of the minimum monitored time required for each burden proportion is plotted against the proportion of the total burden (dotted line, patients C and D, bottom). Patient C developed 50% of his total burden in 11% of the monitored time (black dot, patient C, bottom). Patient D, in contrast, required 40% of the observation time to develop 50% of his burden (black dot, patient D, bottom) as each day contributes less to the total burden because the AF burden is spread over more days. The black diagonal line (patients C and D, bottom) represents a hypothetical uniform AF burden development. The area between the actual (blue or red dotted line) and the uniform hypothetical (solid black diagonal) AF burden development is evaluated as a measure of the temporal aggregation of the AF burden (AF burden density).

![Figure 3](image2.png)

**Figure 3.** Histogram of the patients’ atrial fibrillation (AF) burdens. A total of 174 patients with no documented AF episode during the monitored time were excluded.
The results of such studies comparing continuous versus actual IRM are severely influenced by the choice of time points when the IRM is performed. As depicted in Figure 1, in patient A, a 24-hour HM test performed at time point 1 will be negative and patient A will be regarded as "AF free," whereas the same examination performed some days earlier at time point 2 would have identified AF recurrence.

Although this patient population presents an inhomogeneous collection of patients being monitored continuously with implantable devices, this gives us the opportunity to observe the characteristics of a very wide range of AF burden spectrum. The lower end of the AF burden spectrum is predominantly occupied by patients implanted with a CM device for monitoring after ablation procedures (mean AF burden, 0.05 ± 0.16; median, 0.002; quartile 1, 0.001; quartile 3, 0.01) or patients with paroxysmal, asymptomatic AF, whereas the higher end of the spectrum is predominantly occupied by patients implanted with a pacemaker device for symptomatic bradycardia and long-lasting persistent AF. This wide range of AF burden observed allowed us to evaluate the sensitivity of the IRM strategies and evaluate the AF burden dynamics over a wide range of AF burdens.

**Discussion**

This study provides insights into the sensitivity of various monitoring strategies for the detection of AF recurrence with the use of data from a large patient population monitored with CM devices. Although 2 different monitoring devices were employed, AF detection in both devices was similar. The aim of our study was not to evaluate the AF detection process of the devices, which has been extensively validated previously, but to evaluate the characteristics and dynamics of AF recurrence detection after AF episode registration.
an ablation procedure with the use of a single, postinterventional 30-day HM in our patient population will have a therapeutic success will be overestimated. Similarly, a single, misclassified, AF recurrence will be underdiagnosed, and cautiously because a great proportion of patients will be ecological or interventional therapies when a 24-hour HM is detection of AF recurrence in these patients. The direct detection of AF in almost half of these patients. This low sensitivity becomes even more striking if one considers that the flipping of AF in blocks of AF, a significantly (P<0.0001) higher monitoring frequency is provided. Even in the theoretical scenario of using three 30-day HM tests (for a total of 90 monitored days per year) would have a sensitivity of 82%; however, such a strategy seems unrealistic because it has been shown that complex or prolonged IRM strategies severely affect patient compliance. To obtain the sensitivities of the various IRM strategies presented in this report, a 100% patient compliance is assumed; however, this is rarely the case in everyday clinical practice because even short-duration IRM's pose a significant compromise in the patients' quality of life, and this has been shown to lead to monitoring discontinuation. Roten et al recently showed that even with medium-duration IRM (7-day HM), 42% of patients complained of discomfort and skin irritation, whereas 16% discontinued the monitoring because of this reason. On the other hand, the implantation of leadless CM requires a minor surgical procedure, and the device itself may carry a risk of infection or patient discomfort. Prolonged IRM strategies not only have reduced patient compliance but also offer diminished returns in terms of sensitivity gained per monitored day. The latter problem is depicted in Figure 4. To achieve the same sensitivity (0.63; solid black horizontal line, Figure 4), 30 days of monitoring would have been required with a 30-day HM strategy (one 30-day HM test) versus only 7 days of monitoring with a 24-hour HM strategy (seven 24-hour HM tests). This disproportionate increase in required monitoring time to achieve a certain level of sensitivity between long- and short-duration IRM would certainly have an impact on patient compliance with aggressive IRM strategies.

If the AF burden temporal characteristics are also taken into consideration (Figure 6), the detection of AF recurrence in patients with high-density AF becomes even more problematic. In the case of patients with high AF burden densities, a strategy of six 24-hour HM tests or one 30-day HM test would have been needed to provide a sensitivity of 50% (similar to that of a fair coin). In the low-density group, six 24-hour HM tests or one 30-day HM test would have provided a sensitivity of 0.72 and 0.86, respectively. Thus, patients with high-density AF require an even more aggressive IRM strategy to detect AF recurrence and evaluate results of therapeutic interventions. It becomes clear that with against CM, chance has an immeasurable effect on the results of such studies, and any inferences drawn from these results will be problematic. In the present study, the reconstruction of the rhythm history of every patient allows us to evaluate the true probability of AF recurrence detection of any random IRM of any duration in every patient, and computationally intensive simulation can draw inferences on the sensitivities and success rates of IRM of any duration and frequency.

Figure 4 presents the results of this approach in our patient collective, depicting the sensitivities of the 4 most common IRM durations (24-hour and 7-, 14-, and 30-day HM) at various monitoring frequencies. In our patient population, a monitoring strategy with four 24-hour HM tests would have a sensitivity of only 52%, thus failing to identify recurrence of AF in almost half of these patients. This low sensitivity becomes even more striking if one considers that the flipping of a fair coin would have provided the same sensitivity for the detection of AF recurrence in these patients. The direct implication of these results is that the evaluation of pharmacological or interventional therapies when a 24-hour HM is employed as a monitoring strategy should be interpreted cautiously because a great proportion of patients will be misclassified, AF recurrence will be underdiagnosed, and therapeutic success will be overestimated. Similarly, a single, random 30-day HM in our patient population will have a sensitivity of 63%, and thus the evaluation of the success of an ablation procedure with the use of a single, postinterventional, 30-day HM would have a sensitivity of only 63% (slightly better than the sensitivity the flipping of a fair coin would provide). Even in the theoretical scenario of using three 30-day HM tests (for a total of 90 monitored days per year) would have a sensitivity of 82%; however, such a strategy seems unrealistic because it has been shown that complex or prolonged IRM strategies severely affect patient compliance. To obtain the sensitivities of the various IRM strategies presented in this report, a 100% patient compliance is assumed; however, this is rarely the case in everyday clinical practice because even short-duration IRM's pose a significant compromise in the patients' quality of life, and this has been shown to lead to monitoring discontinuation. Roten et al recently showed that even with medium-duration IRM (7-day HM), 42% of patients complained of discomfort and skin irritation, whereas 16% discontinued the monitoring because of this reason. On the other hand, the implantation of leadless CM requires a minor surgical procedure, and the device itself may carry a risk of infection or patient discomfort. Prolonged IRM strategies not only have reduced patient compliance but also offer diminished returns in terms of sensitivity gained per monitored day. The latter problem is depicted in Figure 4. To achieve the same sensitivity (0.63; solid black horizontal line, Figure 4), 30 days of monitoring would have been required with a 30-day HM strategy (one 30-day HM test) versus only 7 days of monitoring with a 24-hour HM strategy (seven 24-hour HM tests). This disproportionate increase in required monitoring time to achieve a certain level of sensitivity between long- and short-duration IRM would certainly have an impact on patient compliance with aggressive IRM strategies.

![Graph showing sensitivity of various intermittent rhythm monitoring durations at various frequencies in patients with high (>0.5) and low (<0.5) atrial fibrillation (AF) burden density. Both high- and low-density AF groups had similar AF burdens (burden of low-density group: mean, 0.12; median, 0.07; quartile 1, 0.01; quartile 3, 0.18; burden of high-density group: mean, 0.13; median, 0.06; quartile 1, 0.01; quartile 3, 0.19; P=0.36). Low-density AF patients achieve higher AF recurrence detection sensitivity with less frequent monitoring than patients with high-density AF. In patients with high-density AF (blocks of AF), a significantly (P<0.0001) higher monitoring frequency is required to detect recurrence of AF. Only patients with documented AF recurrence were included. HM indicates Holter monitoring.](http://circ1ahajournals1org/Downloaded from)

### Table 2. Linear Regression Analysis of Effect of Burden and Burden Density on the Probability Gain of AF Recurrence Detection of 30-Day and 24-Hour HM vs Continuous Monitoring

<table>
<thead>
<tr>
<th>Factor</th>
<th>Coefficient</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain in AF recurrence detection:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>continuous vs 24-h HM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF burden</td>
<td>−5.76</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>AF burden density</td>
<td>2.61</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Gain in AF recurrence detection:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>continuous vs 30-d HM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF burden</td>
<td>−4.08</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>AF burden density</td>
<td>4.84</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

AF indicates atrial fibrillation; HM, Holter monitoring. To restore normality, the log of the probability gain was used as dependent variable.
intermittent monitoring altogether, a great part or even the majority of patients will be misclassified, and therefore the scientific evaluation of therapeutic interventions for AF becomes problematic.

**Semiquantitative Estimation of AF Burden Based on Negative IRM**

Although IRM may fail to detect AF recurrence in a great proportion of patients and cannot provide the level of information a CM strategy does, results of serial IRM may guide patient management in a semiquantitative way when CM is not possible because of either limited availability, patient compliance, or cost considerations. Such an evaluation is depicted in Figure 5. A series of negative IRM, although may underestimate qualitative AF detection per se, can however guide estimation of the potential AF burden level in patients prone to AF recurrence such as patients after ablation procedures, or patients on rhythm control strategies. In our large population of CM patients, 4 random negative 24-hour HM tests indicate, with a confidence of 90%, an AF burden of up to 23%. A higher number of serial negative IRM tests or negative IRM tests of greater duration can pinpoint the potential AF burden to lower levels. However, it is important to note that even average burdens that low have been shown to be a risk factor for thromboembolic complications.13–16

**AF Characteristics Influencing the Success of IRM**

A major variable driving the sensitivity in detecting AF recurrence is not only the AF burden but the dispersion of it over time. AF density, as a measure of this dispersion, proved to be a major factor that influences the sensitivity of IRM at various frequencies and durations. Although patients A and B of Figure 1 have very similar amounts of AF burden over the same observation time, the probability of AF recurrence detection is vastly different because of the different temporal AF burden dispersion. In patient B (Figure 1), because of the even dispersion of AF burden, prolonged-duration IRM (30-day HM) is not superior to shorter-duration IRM (24-hour HM), the probability of AF recurrence detection of both IRM tests is 1, and the sensitivity of both strategies is 100%. In patient A (Figure 1), however, because of the high AF density, a random 30-day HM test offers only a negligible 3% increase in probability of AF recurrence detection compared with a random 24-hour HM test. In the whole patient cohort (Figure 6), the AF density appears to have a major and statistically significant influence on the sensitivity of AF detection with the use of IRM strategies. In patients with high-density AF, the probabilities of AF recurrence detection and the sensitivities of IRM strategies are significantly inferior to those of patients with same level of AF burden but low AF density (Figure 6).

**Implications for Clinical Outcomes, Patient Management, and Current Knowledge**

Although the level of AF burden has been shown to be a significant risk factor for thromboembolic events.13–16 to the best of our knowledge, no study has taken into consideration the temporal characteristics of AF burden development. It is conceivable that the temporal distribution of AF burden may also play a role in the general thromboembolic risk that patients with AF face. Patients A and B (Figure 1), although spending the same time in AF, may have a different thromboembolic risk because of the different temporal distribution of their AF burden. Larger studies are required to evaluate this hypothesis, and prospective studies are required to evaluate the clinical impact of the combined quantitative and temporal AF characteristics on patient outcome.

The data presented may have some clinical implications. The results of our study depict the major limitations of IRM not only for management of patients but also for evaluation of the results of pharmacological as well as interventional therapies for AF. From a clinical perspective, accurate patient management can only be obtained with confidence with CM. From a scientific perspective, our evaluation until now of the success of pharmacological and interventional procedures based on IRM seems problematic in light of the results presented in this study. The accurate evaluation of therapies for AF in the era of evidence-based medicine mandates the use of CM.

It is conceivable that the widespread use of CM or the results from large studies of CM patients may require us to reevaluate our definitions of success or failure of therapeutic intervention for AF. It may very well be that the current definition of success of a therapeutic intervention only in terms of complete absence of AF recurrence might be inadequate and that a quantitative approach to AF recurrence may in the future require us to critically revisit our current knowledge and reevaluate our therapeutic interventions and
strategies, our results, and our patient management decision-making process. Whether the use of CM to guide clinical management can also improve patient outcomes remains uncertain at this point and must be investigated in large randomized controlled prospective studies.

Limitations

This study was underpowered to investigate the clinical impact of temporal AF characteristics on patient outcomes. We believe that this should be the scope of large prospective studies or registries of CM patients over sufficient follow-up time. AF episodes of patients primarily implanted with a CM device because of atrial tachycardia or atrial flutter have been included in the analysis. In our experience, although patients are typically classified in 1 distinct rhythm type (atrial flutter, AF, or atrial tachycardia), CM reveals that these patients actually tend to experience multiple types of atrial rhythms. Although 2 different devices were employed in the population of this study, previous studies have extensively investigated the accuracy of AF detection of these devices (sensitivity, $\geq 98\%$; specificity, $\geq 91\%$). Additional, we tried to manually inspect all available episodes and, in cases of uncertainty or dispute, we discarded isolated, short ($<5$ minutes) episodes that could represent artifacts.

Conclusion

Patient follow-up with either short- or long-duration IRM is significantly inferior to CM for the detection of AF recurrence. IRM monitoring strategies will not identify AF recurrence in a great proportion of patients at risk. With short-duration, low-frequency IRM strategies, chance has an immeasurable effect on the evaluation of AF recurrence. Thus, results of pharmacological or invasive interventions evaluated with IRM should be considered with caution. Temporal AF characteristics play a major role in AF recurrence detection with the use of IRM strategies and may influence clinical parameters and outcomes in AF patients. For the scientific, evidence-based evaluation of AF treatments, CM should be strongly recommended.

Acknowledgments

We are thankful to Katrin Meyer and Martina Schröder for secretarial support and to Jana Paise, Anja Paap, and Tobias Frin for data documentation support.

Disclosures

Drs Charitos, Stierle, Baldewig, Robinson, and Sievers have no conflict of interest to disclose. Paul D. Ziegler is a full-time employee of Medtronic ($\approx 10,000$ USD) and stockholder of Medtronic ($\approx 10,000$ USD). Dr Hanke discloses modest lecture honoraria ($\approx 10,000$ USD) from Medtronic.

References


Patient follow-up with either short- or long-duration intermittent rhythm monitoring is significantly inferior to that with continuous rhythm monitoring for the detection of atrial fibrillation (AF) recurrence. Intermittent rhythm monitoring strategies fail to identify AF recurrence in a great proportion of patients at risk. Even with feasible, aggressive intermittent monitoring strategies (such as quarterly 24-hour Holter monitoring), AF recurrence will not be identified in a significant proportion of patients. Because of the low sensitivity of intermittent rhythm monitoring, the results of pharmacological or invasive interventions for AF when evaluated with intermittent monitoring should be considered with caution because chance has an immeasurable effect on the detection of AF recurrence, and thus a great proportion of patients will be misclassified, AF recurrence will be underdiagnosed, and therapeutic success will be overestimated. Novel quantitative and temporal AF characteristics (AF burden, AF density) play a major role in AF recurrence detection with the use of intermittent monitoring strategies and may influence clinical parameters and outcomes in AF patients. For accurate patient management, as well as for the scientific, evidence-based evaluation of AF treatments, continuous rhythm monitoring should be strongly recommended.