Atrial Fibrillation Burden Estimates Derived from Intermittent Rhythm Monitoring are Unreliable Estimates of the True Atrial Fibrillation Burden

EFSTRATIOS I. CHARITOS, M.D.,* PAUL D. ZIEGLER, M.S.,† ULRICH STIERLE, M.D.,* DEREK R. ROBINSON, M.A., M.Sc., D.Phil., C.S.T.A.T.,‡ BERNHARD GRAF, M.D.,* HANS-HINRICH SIEVERS, M.D.,* and THORSTEN HANKE, M.D.*

From the *Department of Cardiac and Thoracic Vascular Surgery, University of Luebeck, Luebeck, Germany; †Medtronic Inc., Minneapolis, Minnesota; and ‡Department of Mathematics, School of Mathematical and Physical Sciences, University of Sussex, Brighton, UK

Background: *Estimates of atrial fibrillation (AF) burden (AFB) derived from intermittent rhythm monitoring (IRM) are increasingly being used as an outcome measure after therapeutic interventions; however, their accuracy has never been validated. The aim of this study was to compare IRM-derived AFB estimates to the true AFB as measured by implantable continuous monitoring (CM) devices.*

Methods: *Rhythm histories from 647 patients (mean AFB: 12 ± 22%; 687 patient·years) with CM devices were analyzed. IRM of various frequencies and durations were simulated and the obtained IRM-derived AF burdens were compared to the true AFB measured by CM.*

Results: *The relative error of the IRM burden estimates was dependent on the IRM length ($P < 0.001$), frequency of IRM ($P < 0.001$), the true AFB ($P < 0.001$), and its temporal aggregation (AF density, $P < 0.001$). In paroxysmal AF patients, the relative error even with aggressive IRM strategies was >80% of the true AFB. The relative error decreased with higher true AF burdens, lower AF densities, and higher IRM frequency or duration ($P < 0.001$). However, even in patients with high AF burdens and/or low AF densities, IRM estimates of AFB significantly deviated from the true AFB (relative error >20%, $P < 0.001$) and resulted in a substantial measurement error.*

Conclusion: *IRM-derived AFB estimates are unreliable estimators of the true AFB. Particularly for paroxysmal AF patients, IRM-derived AFB estimates should not be used to evaluate outcomes after AF interventions. (PACE 2014; 00:1–9)*

atrial fibrillation, arrhythmia, monitoring, clinical trial design

Introduction

Recent studies have shown that patient follow-up with intermittent rhythm monitoring (IRM) fails to detect atrial fibrillation (AF) recurrence and introduces significant bias, overestimates the success rates, and distorts the scientific evaluation of therapeutic interventions.[1–3] Added to this, the outcome of "time-to-first-AF recurrence" has been shown to be an inappropriate end point for clinical trials[4,5] due to the tendency of AF to recur in temporal clusters. Intracardiac[6] and implantable, leadless, subcutaneous monitoring devices[7] are now capable of continuously monitoring AF episodes. This has led to the concept of AF burden (AFB) defined as the proportion of the observed time that a patient is in AF.[7–9] Both intracardiac and subcutaneous monitoring devices reliably quantify AFB with ≥98.5% accuracy.[7,10] In contrast to other clinical end points (symptoms, hospitalizations, or rare clinical complications of AF such as stroke, bleeding, or mortality), AFB can be measured on a continuous scale and in every patient, especially in the setting of a clinical trial comparing AF treatments. Consequently, drawing inferences on the success of a therapeutic intervention based on a quantitative outcome that can be measured in every patient may require

Sources of funding: The computational resources for this work were kindly provided by AWS WA, USA (Education Coursework Grant Award: *EDU_Charitos_ULuebeckResearch_Fall2012*). The grant donor had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. No other sources of funding to disclose.
Clinical Trial Registration URL: http://www.clinicaltrials.gov. Unique identifier: NCT00806689.
Disclosures: Drs. Charitos, Stierle, Graf, Robinson, and Sievers have no conflict of interest to disclose. Paul D. Ziegler is a full-time employee of Medtronic (>10,000 USD) and stockholder of Medtronic (>10,000 USD). Dr. Hanke discloses modest lecture honoraria (<10,000 USD) from Medtronic.

Address for reprints: Efstratios I. Charitos, M.D., Department of Cardiac and Thoracic Vascular Surgery, University of Luebeck, Ratzeburger Allee 160, 23568 Luebeck, Germany. Fax: 49-451-500-2051; e-mail: efstratios.charitos@gmail.com

Received November 27, 2013; revised January 3, 2014; accepted February 14, 2014.

doi: 10.1111/pace.12389

shorter studies with smaller sample sizes and therefore less costly clinical trials.

Primarily due to the associated costs and invasiveness, patient follow-up using solely continuous monitoring (CM)—both in the clinical and investigational settings—may not be feasible. Recently, several studies have been published in which AFB is estimated with longer duration (7 days, 14 days) IRM performed several times in every patient throughout the course of the study.[11–13] This IRM-derived AFB estimate (IRM-AFB) is assumed to represent the patients' true AFB and is then used as an end point to evaluate the AF status of the patients and the success of therapeutic interventions. Although the use of IRM-AFB estimates seems logical, to our knowledge the reliability of these estimates has never been validated or compared to the true AFB.

The aim of this study is to evaluate how IRM-AFB estimates of various IRM frequencies and durations compare to the true AFB, as measured in a large population of patients monitored with implantable CM devices.

Methods

Population Characteristics, Cardiac Rhythm Reconstruction, and IRM Simulation

Data acquired from 647 patients monitored with a CM device (Reveal XT 9529, n = 73; AT500 pacemaker, n = 574; Medtronic, Inc., Minneapolis, MN, USA) were analyzed. Demographics and detailed patient characteristics have been reported previously[1] and are presented in detail in the online Supporting information (part A). The mean true AFB was 0.12 ± 0.22 and the mean follow-up was 1.1 ± 0.4 years (range: 0.1–3.7 years; 687 patient years). All patients provided informed consent for the data collection and use. The study was approved by the local ethics committee (ClinicalTrials.gov ID: NCT00806689).

In every patient, the complete rhythm history was reconstructed. Thereafter, we used computationally intensive simulation methods to simulate IRMs of various durations (1, 2 , 3,. . . ,30 days) and frequencies (n = 1, 2, 3,. . . ,12) in every patient and evaluated the IRM-AFB obtained by IRM strategies in all patients. The benefits of using a computational approach in contrast to a direct use of real IRMs is that a computational approach allows us to perform an unlimited amount of "virtual" IRM strategies of any duration, frequency, and level of compliance in all patients and evaluate the obtained IRM-AFB estimates against the true values. In short, in a conventional study design, the choice of time when the IRM takes place has an enormous influence on the results.[1] Because AF is intermittent, IRM on given day(s) may detect a certain amount of AFB; however, if the same examination happens to take place several days earlier or later, this could change the results and lead to a different AFB estimation. Because in the most conventional studies, only a limited number of IRM tests were used, chance has an immeasurable effect on the results of such studies, and any inferences on AFB drawn from these results will be problematic. In this study, we reconstructed the rhythm history of every patient and used computationally intensive simulation to simulate in every patient all possible IRM strategies of various durations and frequencies and to draw inferences on the obtained distribution of the IRM AFB and true measurement error of IRM AFB versus the true AFB for IRM of any duration and frequency. Technical details of the simulation procedure are presented in the online Supporting information (part B).

IRM-AFB was defined as the proportion of the total IRM monitored time that the patient was in AF. The AF density, as described previously,[1] was evaluated as a quantitative measure of the AFB temporal aggregation and was calculated as an index consisting of values between 0 (AFB evenly spread over the observation time) and 1 (maximum possible AFB aggregation, i.e., "one block of AF").[1] Details on the calculation of the AF density are presented in the online Supporting information (part C).

Because significant evidence exists that longer duration IRM leads to reduced patient compliance for various reasons,[7,14–17] thus leading to data loss, we simulated various levels of compliance to investigate how a reduction in patient compliance may affect the measurement error of IRM-AFB estimates. Compliance was defined as the proportion of actual monitored days to the number of days initially planned for monitoring. Loss of data due to reduced compliance was assumed to be random. Compliance was investigated at the 90%, 75%, and 50% levels.

Estimation of IRM-AFB Measurement Error

For the measurement of the error between the true AFB estimates and the IRM-AFB estimates, the relative mean absolute deviation (RAD) was calculated in all patients and for all IRM strategies (IRM frequencies and durations) and levels of compliance. Technical details on the calculation of RAD are presented in the online Supporting information (part D). RAD provides a robust measure of the IRM-AFB error as a percentage of the true value and is presented in all analyses included in this work. For example, a relative error of 50% of a true burden 0.20 would indicate that the *average expected error* of the IRM-AFB estimate is ±0.1.

Statistical Analyses

The influence of AFB and density, and IRM duration and frequency on RAD, was examined using second-order response surface regression models. Details on the modeling procedure are presented in the online Supporting information (part E). Separate models were used for the most commonly used IRM durations (24 hours, 7 days, 14 days, and 30 days). All statistical analyses and simulations were performed with R version 2.15.2 (R Development Core Team 2012). The P values of two-sided tests at a significance level of 0.05 are reported.

Results

Determinants of IRM-AFB Error

Regression models showed that the determinants of RAD were the AF characteristics (AFB, $P < 0.001$; AF density, $P < 0.001$) as well as the duration of the IRM ($P < 0.001$) and the IRM frequency (number of IRM/year, $P < 0.001$). The detailed regression models and results, together with the effect sizes and interactions of the AF and IRM characteristics on the error of IRM-AFB (response surface models) are presented in online Table 1 (Supporting information, part F). These results are visualized more clearly in Figure 1. For any AF density, RAD increases at lower AFB. Higher AF densities result in significantly higher RAD. The IRM duration had a significant effect on RAD with increasing IRM duration leading to decreased RAD. This effect was more pronounced at low densities (Fig. 1, left) and became less important at high AF densities (Fig. 1, right). The relative error of the most commonly used IRM strategies (24-hour, 7-day, and 30-day IRM) to estimate AFB is shown in Figure 2. Our analyses presented in Figures 1 and 2 as well as in online Table 1 (Supporting information, part F) indicate that IRM-AFB estimates result in a significant measurement error at all levels of AFB and density and for all IRM strategies. This measurement error becomes particularly concerning at true burdens <0.4 for which the RAD exceeds values of 80% (Figs. 1 and 2).

Averaging the IRM burden estimates obtained from multiple IRM of the same duration (IRM frequency, $P < 0.001$, online Table 1 [Supporting information, part F]; dotted lines in Fig. 1) at different time points in each patient resulted in significantly reduced variability and RAD of the IRM-AFB estimates. The effect of averaging IRM-AFB was independent of the AF and IRM characteristics (online Table 1 [Supporting information, part F], Fig. 1). The reduction of RAD provided by averaging of IRM-AFB was more pronounced at lower IRM durations (24-hour IRM, Fig. 1).

Effect of Patient Compliance on IRM-AFB Error

Patient compliance had a statistically significant influence on the measurement error of IRM burden estimates (RAD) and this influence was independent from the AF or IRM characteristics. Regression models showed that reduced compliance led to a significant *additional* increase in RAD for every type of AF and IRM characteristic.

For 7-day IRM, compliance levels of 90%, 75%, and 50% led to an additional average increase of RAD by 6.1% (95% confidence interval [CI]: 5.1–7.0%; $P < 0.001$), 16.7% (95% CI: 15.3–17.9%; $P < 0.001$), and 24.8% (95% CI: 23.3–26.2%; $P < 0.001$), respectively.

For 14-day IRM, compliance levels of 90%, 75%, and 50% led to an additional average increase of RAD by 4.4% (95% CI: 4.2–4.6%; $P < 0.001$), 10.1% (95% CI: 9.7–10.5%; $P < 0.001$), and 24.4% (95% CI: 22.9–25.9%; $P < 0.001$), respectively.

For 30-day IRM, compliance levels of 90%, 75%, and 50% led to an additional average increase of RAD by 3.1% (95% CI: 2.7–3.3%; $P < 0.001$), 9.0% (95% CI: 8.5–9.6%; $P < 0.001$), and 19.2% (95% CI: 18.2–20.4%; $P < 0.001$), respectively.

Discussion

Magnitude of IRM-AFB Error

Our results indicate a disconcerting unreliability of IRM-AFB estimates. Regardless of the IRM and AF characteristics, the relative error of the IRM-AFB estimates greatly exceeded the level of 10% of the true AFB that might be considered acceptable for a diagnostic procedure (Fig. 1, black dotted line). For AF burdens <0.6, the error of the IRM-AFB estimates is larger than 50% of the true value. For AF burdens <0.3, which is a population that has recently been used for the evaluation of the clinical effectiveness of novel AF treatments,[11] the error of IRM-AFB estimates exceeds 100% of the true AFB (Figs. 1 and 2). To the best of our knowledge and research, we are unaware of any other diagnostic procedure in medicine entailing such a high amount of measurement error.

This amount of uncertainty can be visualized in Figure 3. Two patients with the same amount of AFB but different temporal aggregation (AF density) are presented on the left, while the results of 10^5 simulations of the average of three 7-day IRM and the probability of obtaining a certain IRM-AFB are displayed on the right. In patient A, the relative error of the IRM-AFB obtained with this strategy is 95% of the true AFB. Although the true AFB is 0.2, almost 40% of all simulated IRMs failed to detect any AF, whereas 28% of all IRM

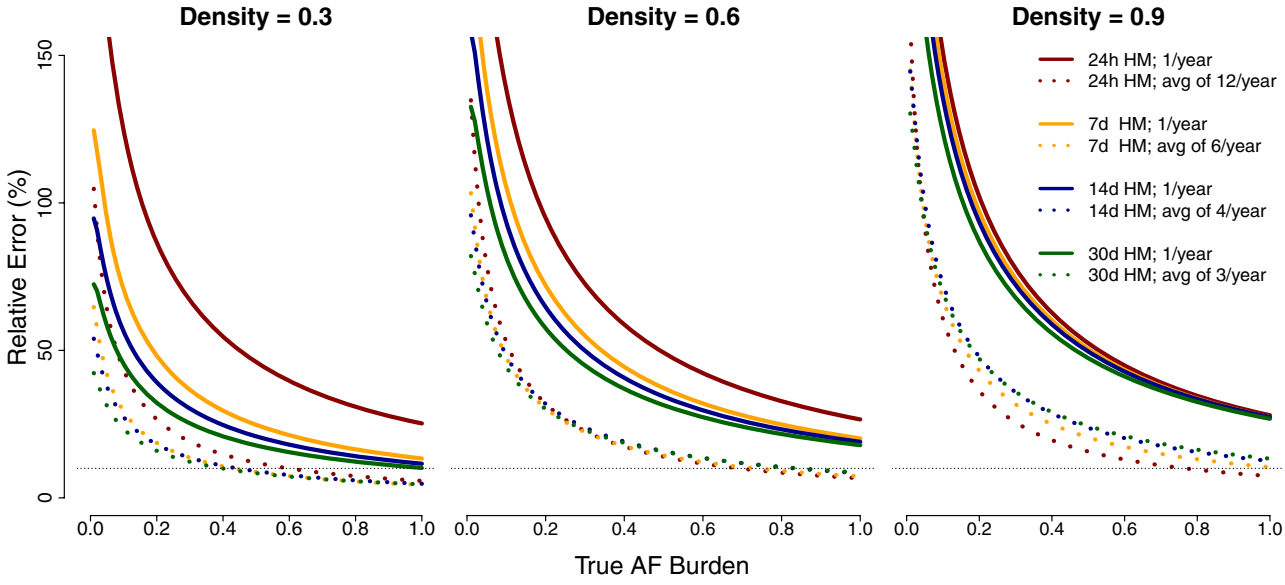

Figure 1. *The relative error (relative mean absolute deviation) of IRM-AFB estimates as a function of the patient's true AFB at three different levels of AF density. The relative error of the AF burdens (point estimates) obtained by averaging multiple IRMs at different times is also displayed (dotted lines). The dotted black horizontal lines depict a relative error level of 10% which may be considered acceptable in diagnostic procedures. AF = atrial fibrillation; AFB = atrial fibrillation burden; IRM = intermittent rhythm monitoring.*

simulations estimated a burden of 0.35%, and 8% of simulations estimated a burden of 0.65%. With this IRM strategy, the probability of obtaining an IRM-AFB that falls within 10% of the true value (blue bars of histogram) is only 1%. In contrast, patient B (whose AFB is more evenly spread across the observation time and therefore has a lower AF density) has a relative error of 24%. Nevertheless, even for this patient, the probability of obtaining an IRM-AFB to within 10% of the true value (blue bars of histogram) is only 28%. However, it is important to note that since the AF characteristics in the setting of a clinical trial or individual patient follow-up are unknown *a priori*, the amount of uncertainty that IRM-AFB estimation entails in every individual patient *prospectively* is inestimable and indeterminable.

Factor Influencing the Measurement Error of the IRM-AFB Estimates

The factors affecting the measurement error of the IRM-AFB estimates were the AF and IRM characteristics (Figs. 1 and 2, online Table 1 [Supporting information, part F]). The IRM-AFB measurement error significantly increases with lower burdens, higher AF densities, and shorter IRM durations (online Table 1 [Supporting information, part F]).

Averaging the AFB estimates obtained from multiple IRM (increase in IRM frequency, online Table 1 [Supporting information, part F]; Fig. 1) throughout the patient observation period resulted in significantly decreased IRM-AFB variability and significant RAD reduction (online Table 1 [Supporting information, part F], Fig. 1). On average, and in the setting of a clinical trial with a large number of patients, this approach would significantly improve the reliability of the IRM-AFB. However, this would result in two distinct disadvantages: first, this approach would reduce a longitudinal study of AFB development to a point estimate comparison, since several IRM-AFB would be averaged to obtain a single, albeit more reliable, estimate. Second, this approach would result in an unacceptably high relative error for a diagnostic procedure (relative error >50%, Fig. 1).

Another factor, which contributes to the measurement error of AFB with IRM, is the issue of patient noncompliance. Previously published simulation studies of IRM for the detection of AF recurrence have conservatively assumed 100% compliance with the scheduled monitoring strategies.[1–3,18–20] In reality, however, the actual IRM compliance is much lower for a variety of reasons. Reported causes of patient noncompliance with scheduled monitoring include skin irritation, interference with showering or exercise, and feelings of self-consciousness when wearing the monitoring equipment in public.[21] Factors unrelated to the patient (such as failure of the batteries, recording media, or adhesive patches) also contribute to loss of data with IRM. A recent

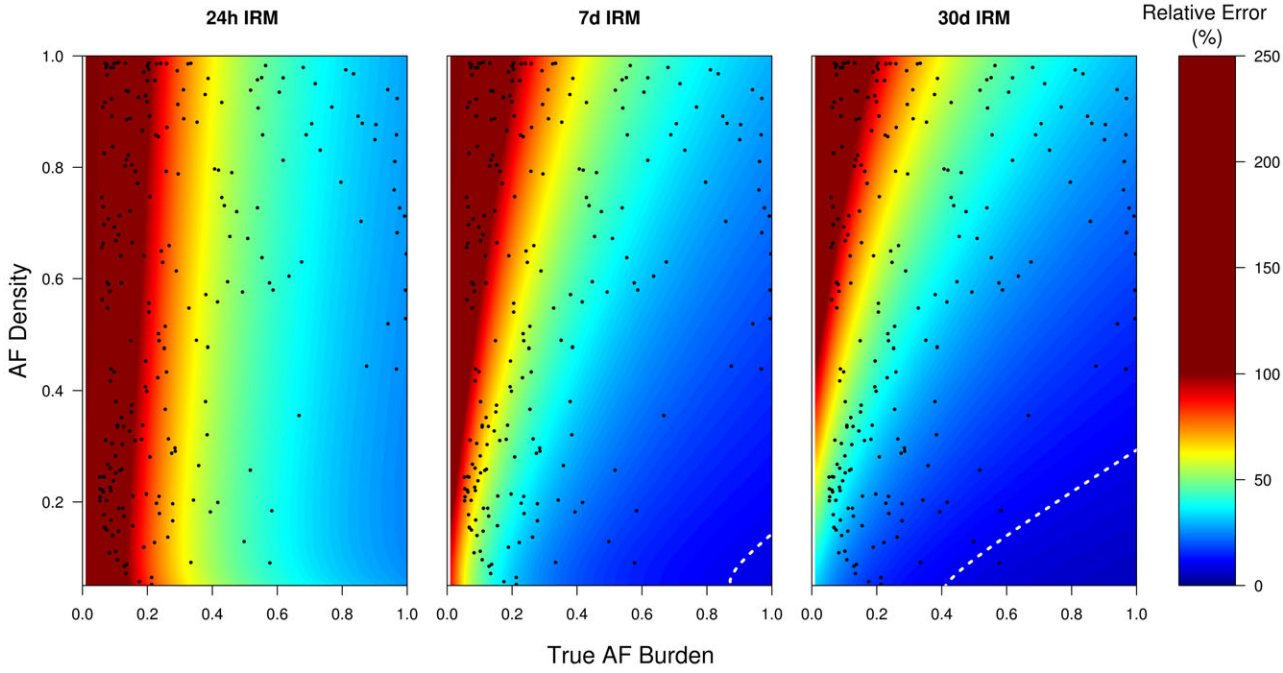

Figure 2. *The relative error (relative mean absolute deviation) of the AFB estimation derived from the most commonly used IRM. The relative error (relative mean absolute deviation) of the AFB estimation derived from the most commonly used IRM (24 hours, 7 days, and 30 days) as a function of the patient's true AFB and AF density. The dotted white lines delineate a relative error level of 10%. The black points denote the AFB and AF density combinations of our patient population. Abbreviations as in Figure 1.*

study with an external arrhythmia monitoring patch reported that the device fell off in 22% of patients and resulted in a mean wear time of 7.9 ± 1.8 days instead of the planned 14 days.[22] In a pilot trial of 40 patients with cryptogenic stroke or high-risk transient ischemic attack, patients were randomly assigned to mobile cardiac outpatient telemetry monitoring for 21 days or to routine follow-up.[17] Kamel et al. reported that the patients in the aggressive monitoring group only wore the monitors for 64% of the assigned days and that 25% of patients were not compliant at all with the scheduled monitoring.[17] In another study, which used the same monitoring technology in 19 patients who recently underwent catheter ablation for AF, only 53% of patients complied with the scheduled monitoring.[16] In our study, we analyzed the effect of 90%, 75%, and 50% compliance with scheduled IRM. Not surprisingly, poor compliance contributed to an *additional* AFB measurement relative error of up to 25%. Although a given rate of noncompliance had a greater impact on less-rigorous IRM strategies, it is important to recognize that the rate of noncompliance (whether due to patient or technical factors) is likely to be higher in the more rigorous IRM strategies.

Implications for Causal Inferences on the Success of Therapeutic Interventions in the Presence of High Uncertainty

On an individual patient level, the high variability of IRM-AFB estimates makes causal inferences regarding changes in observed IRM-AFB estimates particularly precarious. This can be illustrated in Figure 4. This patient has been continuously monitored for approximately 2 years, and a therapeutic intervention (change of medication therapy) took place in the middle of the observation period (green line, Fig. 4, left panel). In this patient, the period prior to the intervention (Period A, Fig. 4, left panel) happened to have the same AFB and density as the postintervention monitored period (Period B, Fig. 4, left panel). With the use of CM and cardiac rhythm history reconstruction, we can conclude with confidence that the intervention did not result in a significant AFB reduction. In this patient, the relative error of IRM-AFB with averaging the burden of three 7-day IRMs (in each Period A and B) is 43.1%, and the probability of obtaining an IRM-AFB within 10% (blue columns of histogram, Fig. 4, right) of the true burden (dotted black line, Fig. 4, right) is

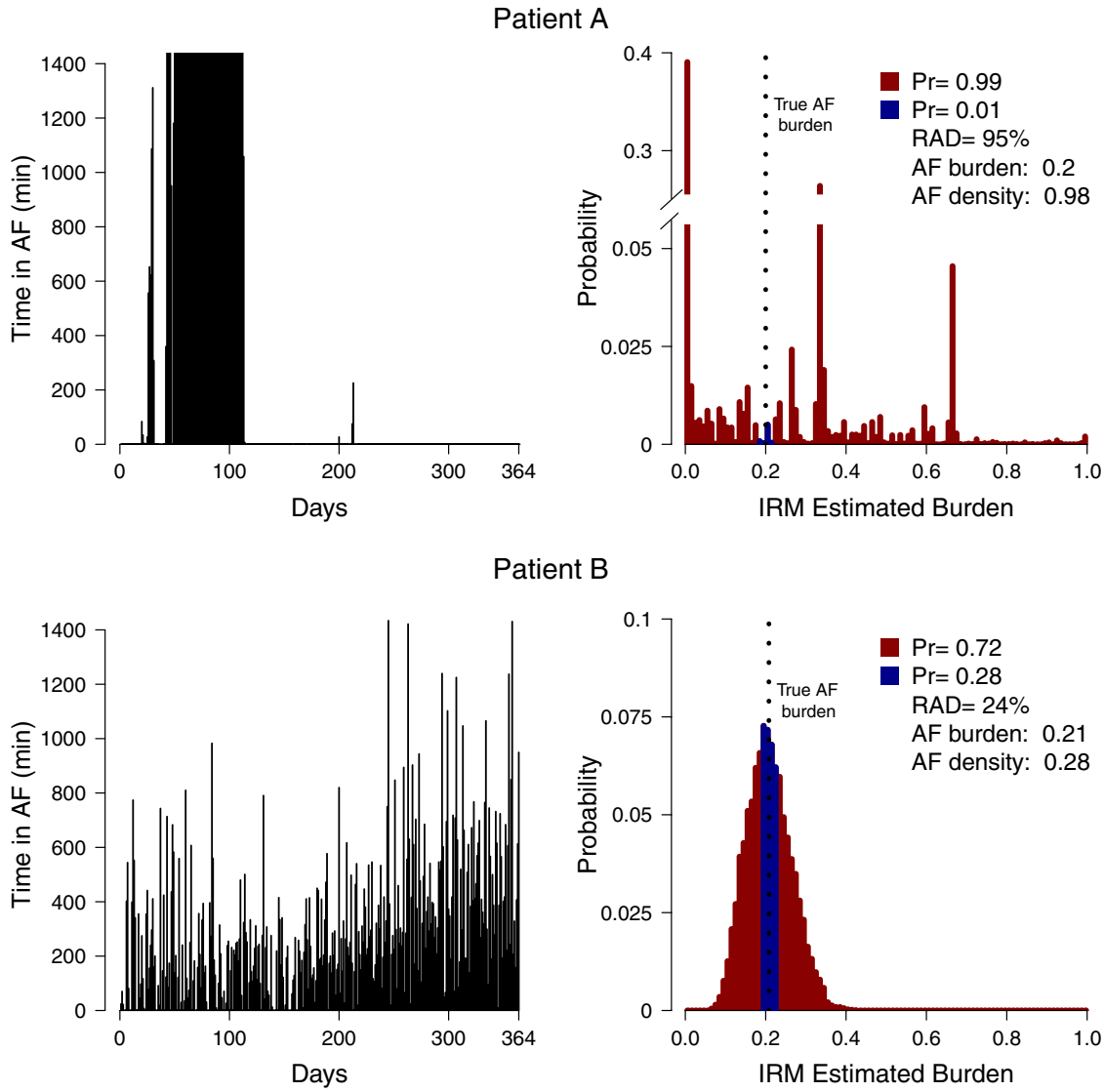

Figure 3. *Upper panel: cardiac rhythm reconstruction (left) and IRM AFB distribution (average of three 7-day IRM per year, right) of a patient with high-density AF. Lower panel: cardiac rhythm reconstruction (left) and IRM AFB distribution (average of three 7-day IRM per year, right) of a patient with a similar AFB to patient A but low-density AF. Abbreviations as in Figure 1.*

only 13%. Since Periods A and B happen to have almost the same burden and density, the IRM-AFB distribution that would be obtained with averaging three 7-day IRMs in each period is the same and is displayed in Figure 4, right panel. In this probability distribution, the probability of obtaining an IRM-AFB <0.09 (burden range C, Fig. 4, right) is the same as the probability of obtaining an IRM-AFB >0.35 (Pr = 0.2, burden range D, Fig. 4, right).

Assuming that an IRM strategy performed in Period A resulted in an IRM AFB within burden range D (AFB > 0.35) and that the follow-up IRM strategy performed in Period B resulted in an IRM AFB within burden range C (AFB < 0.09), the logical implication would be that the therapeutic intervention was successful and led to a true decrease in the patient's burden. Conversely, if an IRM strategy performed in Period A resulted in an IRM AFB within the burden range C (AFB < 0.09) and the follow-up IRM strategy performed in Period B resulted in an IRM AFB within the burden range D (AFB > 0.35), the logical implication would be that the therapeutic intervention was unsuccessful and that the patient's AF status and burden

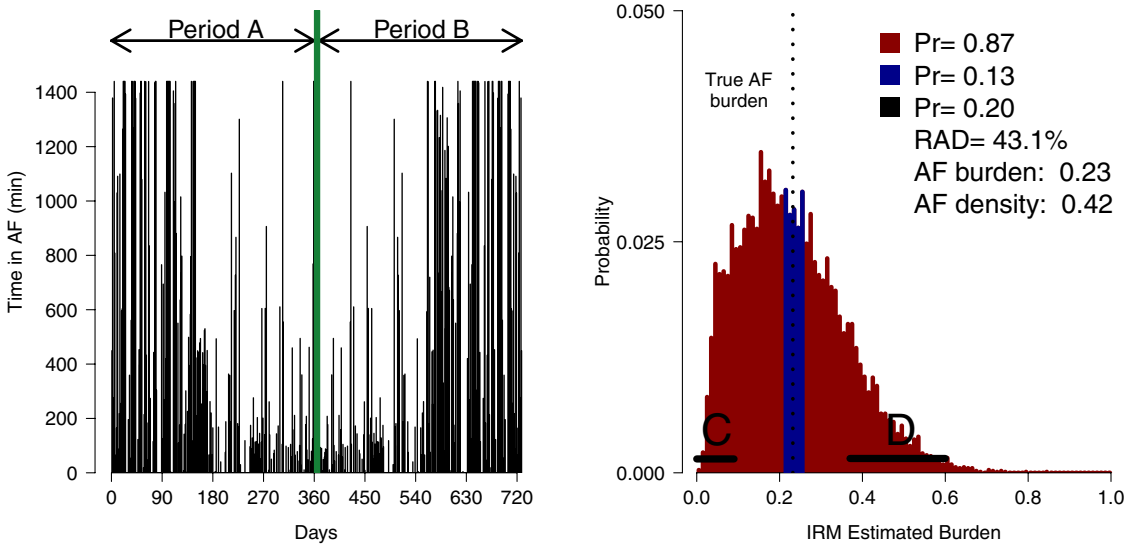

Figure 4. *A patient is continuously monitored for approximately 2 years. A therapeutic intervention occurred in the middle of the observation period (green solid vertical line, left panel). The period prior to the intervention (Period A) happens to have the same AFB and density as the postintervention period (Period B). Continuous monitoring and cardiac rhythm history thus indicates that the intervention did not result in a significant AFB reduction. The variability of the IRM-AFB estimates can severely distort the evaluation of therapeutic interventions. Abbreviations as in Figure 1.*

deteriorated. The reality, however, as revealed with CM, is that the intervention resulted in no change in the patient's AFB.

It is striking to note that the probability of obtaining either of the above two very different interpretations and conclusions is the same and is greater than the probability of obtaining an IRM-AFB estimate within 10% of the true AFB (Pr = 13%; blue columns of histogram, Fig. 4, right). The variability of the IRM-AFB estimates can therefore severely distort the scientific, evidence-based evaluation of therapeutic interventions. The different causal inferences that could have been obtained with IRM-AFB estimates should be attributed to the high error, uncertainty, and randomness entailed in the IRM-AFB estimates and not to the success or failure of the therapeutic intervention. This amount of uncertainty may have implications for individual patient follow-up. Patient follow-up in the clinical setting is per definition a longitudinal process and as such the estimation of AFB for any interval using IRM-AFB estimates will lead to significant measurement error and precarious causal inferences. An additional disadvantage of IRM-AFB estimation is that any information about the progression of AF as a disease will probably be lost either due to the high variability of the IRM-AFB as an estimator, or in the process of averaging the results of IRM-AFB measurements in order to obtain a more reliable estimate.

Implications for the Design of Clinical Trials

The variability of an estimator—in this case the IRM-AFB—is fundamental for establishing the statistical significance of a comparison between samples. By extension, due to the high variability of IRM-AFB, studies using IRM-AFB will inevitably have a high probability of Type II error (failing to reject a false null hypothesis ≈ failing to statistically establish a difference that is otherwise true). This can be visualized in Figure 5 which shows two samples of simulated patients with significantly different true AFB (sample A: 0.35, sample B: 0.2, solid horizontal lines, Fig. 5) and with various AF densities. A strategy of one 7-day IRM strategy would fail to show statistical significance for the difference in the AFB of these two groups with a probability of 72.2% (statistical power of only 27.8%) due to the high variability and wide overlap of the IRM-AFB distribution in these populations (thin red and blue lines, Fig. 5). It is noteworthy that there are several studies that have avoided the use of IRM-AFB and instead have employed CM to document AFB reduction between groups.[3,23–26]

Fundamental theorems in probability theory (law of large numbers, central limit theorem) applied in our results indicate that although IRM-AFB estimates are imprecise, these estimates will tend to the population's true AFB given that the sample size is *sufficiently large*. However, due to the increased error and variance of the

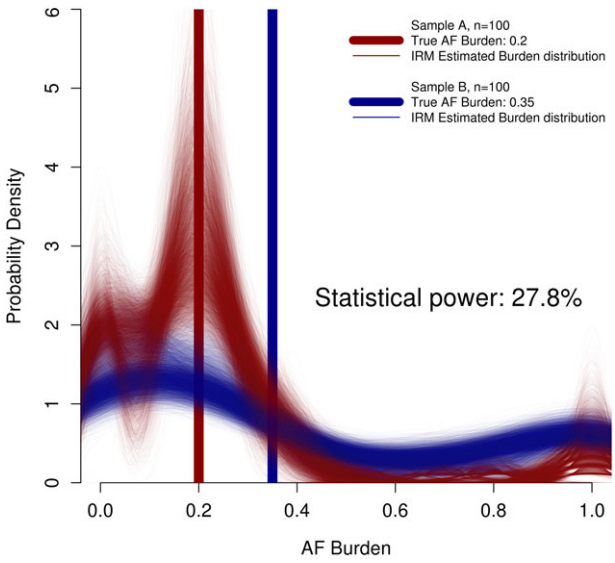

Figure 5. *Simulation of two samples of 100 patients, each having different true AFB (solid vertical lines, Sample A: 0.2, Sample B: 0.35) with various AF densities. The simulated IRM-AFB distributions (7,500 simulations) in these two samples are presented with the horizontal thin lines. Due to the high variability and the wide overlap of the IRM-AFB estimates in these two samples, the probability of Type II error (failing to reject a false null hypothesis≈failing to statistically establish a difference that is true) is 72.2% and the statistical power to detect a difference that truly exists is only 27.8%. Abbreviations as in Figure 1.*

IRM-AFB and additional variance due to compliance or potential data loss issues, studies employing IRM-AFB estimates will inevitably have a reduced statistical power and would require an increased number of patients in order to detect a difference in AFB that otherwise would be true. For example, in our patient population, the use of 7-day IRM-AFB would reduce the power to detect a 40% effect size (decrease in true AFB from 0.25 to 0.1) by up to 50% in a two-arm study with a sample size of 100 patients per arm.

Study Limitations

Our methodology does not take into consideration patient symptoms, which may help guide AF follow-up. Numerous studies have shown that symptoms have a low sensitivity and specificity,[27–30] which may or may not lead to better and more reliable AF recurrence detection.[29,31] There is significant evidence in the literature of a disconnection between symptoms and AF recurrence: symptoms perceived to be related to AF often are not associated with an atrial tachyarrhythmia and the vast majority of AF episodes are asymptomatic.[27] Therefore, for the scientific evaluation of AF treatments even with the presence of symptoms (and because of their low sensitivity and specificity), AF recurrence should still be electrocardiographically documented. In addition, recent evidence shows that the ratio of asymptomatic to symptomatic AF episodes increases after invasive AF treatments.[23] Therefore, although a limitation, we believe that not allowing for patient symptoms does not limit the validity and applicability of our findings. The majority of our patient population had low AFB (mean burden 0.12 ± 0.22, median 0.11, third quartile 0.13). Although our population does include patients with high burden, the results of our models might be less accurate in these patients and might not reflect accurately the IRM-AFB error at very high burdens (>0.8). Although this study is a simulation study describing the error that is expected to occur when IRM-AFB estimates are used for patient management or as end points for clinical trials, a clinical trial is still necessary to clarify the clinical meaning of misclassification of AF recurrence and AFB.

Conclusions

We provide evidence that IRM-AFB estimates are unreliable estimators of the true AFB with high variability, low accuracy, low precision, and high uncertainty. The use of IRM-AFB estimates leads to reduced statistical power of studies investigating the impact of AF treatments on quantitative AF status and to an increased probability of Type II error. Results of trials using IRM-AFB estimates should be interpreted with caution, especially if these trials fail to show evidence of AFB differences for subgroups or after therapeutic interventions.

References

1. Charitos EI, Stierle U, Ziegler PD, Baldewig M, Robinson DR, Sievers H-H, Hanke T. 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. Circulation 2012; 126:806–814.
2. Ziegler PD, Koehler JL, Mehra R. Comparison of continuous versus intermittent monitoring of atrial arrhythmias. Heart Rhythm 2006; 3:1445–1452.
3. Martinek M, Aichinger J, Nesser H-J, Ziegler PD, Purerfellner H. New insights into long-term follow-up of atrial fibrillation ablation: Full disclosure by an implantable pacemaker device. J Cardiovasc Electrophysiol 2007; 18:818–823.
4. Gillis AM, Rose MS. Temporal patterns of paroxysmal atrial fibrillation following DDDR pacemaker implantation. Am J Cardiol 2000; 85:1445–1450.
5. Ricci R, Santini M, Padeletti L, Boriani G, Capucci A, Botto G, Gulizia M, et al. Atrial tachyarrhythmia recurrence temporal

patterns in bradycardia patients implanted with antitachycardia pacemakers. J Cardiovasc Electrophysiol 2004; 15:44–51.
6. Purerfellner H, Gillis AM, Holbrook R, Hettrick DA. Accuracy of atrial tachyarrhythmia detection in implantable devices with arrhythmia therapies. Pacing Clin Electrophysiol 2004; 27:983–992.
7. Hindricks G, Pokushalov E, Urban L, Taborsky M, Kuck K-H, Lebedev D, Rieger G, et al. Performance of a new leadless implantable cardiac monitor in detecting and quantifying atrial fibrillation: Results of the XPECT trial. Circ Arrhythm Electrophysiol 2010; 3:141–147.
8. Shah D. Atrial fibrillation burden: A "hard" indicator of therapeutic efficacy and a prognostic marker to boot? Eur Heart J 2008; 29:964–965.
9. Veasey RA, Segal OR, Large JK, Lewis ME, Trivedi UH, Cohen AS, Hyde JA, et al. The efficacy of intraoperative atrial radiofrequency ablation for atrial fibrillation during concomitant cardiac surgery—The Surgical Atrial Fibrillation Suppression (SAFS) Study. J Interv Card Electrophysiol 2011; 32:29–35.
10. Passman RS, Weinberg KM, Freher M, Denes P, Schaechter A, Goldberger JJ, Kadish AH. Accuracy of mode switch algorithms for detection of atrial tachyarrhythmias. J Cardiovasc Electrophysiol 2004; 15:773–777.
11. Cosedis Nielsen J, Johannessen A, Raatikainen P, Hindricks G, Walfridsson H, Kongstad O, Pehrson S, et al. Radiofrequency ablation as initial therapy in paroxysmal atrial fibrillation. N Engl J Med 2012; 367:1587–1595.
12. Winkle RA, Mead RH, Engel G, Kong MH, Patrawala RA. Atrial arrhythmia burden on long-term monitoring in asymptomatic patients late after atrial fibrillation ablation. Am J Cardiol 2012; 110:840–844.
13. Pontoppidan J, Nielsen JC, Poulsen SH, Hansen PS. Symptomatic and asymptomatic atrial fibrillation after pulmonary vein ablation and the impact on quality of life. Pacing Clin Electrophysiol 2009; 32:717–726.
14. Roten L, Schilling M, Häberlin A, Seiler J, Schwick NG, Fuhrer J, Delacrétaz E, et al. Is 7-day event triggered ECG recording equivalent to 7-day Holter ECG recording for atrial fibrillation screening? Heart 2012; 98:645–649.
15. Mittal S, Movsowitz C, Steinberg JS. Ambulatory external electrocardiographic monitoring: Focus on atrial fibrillation. J Am Coll Cardiol 2011; 58:1741–1749.
16. Vasamreddy CR, Dalal D, Dong J, Cheng A, Spragg D, Lamiy SZ, Meininger G, et al. Symptomatic and asymptomatic atrial fibrillation in patients undergoing radiofrequency catheter ablation. J Cardiovasc Electrophysiol 2006; 17:134–139.
17. Kamel H, Navi BB, Elijovich L, Josephson SA, Yee AH, Fung G, Johnston SC, et al. Pilot randomized trial of outpatient cardiac monitoring after cryptogenic stroke. Stroke 2013; 44:528–530.
18. Ziegler PD, Glotzer TV, Daoud EG, Wyse DG, Singer DE, Ezekowitz MD, Koehler JL, et al. Incidence of newly detected atrial arrhythmias via implantable devices in patients with a history of thromboembolic events. Stroke 2010; 41:256–260.
19. Ziegler PD, Glotzer TV, Daoud EG, Singer DE, Ezekowitz MD, Hoyt RH, Koehler JL, et al. Detection of previously undiagnosed atrial fibrillation in patients with stroke risk factors and usefulness of continuous monitoring in primary stroke prevention. Am J Cardiol 2012; 110:1309–1314.
20. Ziegler PD, Koehler JL, Verma A. Continuous versus intermittent monitoring of ventricular rate in patients with permanent atrial fibrillation. Pacing Clin Electrophysiol 2012; 35:598–604.
21. Henry L, Ad N. Long-term monitoring for patients after surgical ablation of atrial fibrillation: Are all devices the same? Innovations (Phila) 2010; 5:259–264.
22. Rosenberg MA, Samuel M, Thosani A, Zimetbaum PJ. Use of a noninvasive continuous monitoring device in the management of atrial fibrillation: A pilot study. Pacing Clin Electrophysiol 2013; 36:328–333.
23. Verma A, Champagne J, Sapp J, Essebag V, Novak P, Skanes A, Morillo CA, et al. Discerning the incidence of symptomatic and asymptomatic episodes of atrial fibrillation before and after catheter ablation (DISCERN AF): A prospective, multicenter study. JAMA Intern Med 2013; 173:149–156.
24. Pürerfellner H, Urban L, de Weerd G, Ruiter J, Brandt J, Havlicek A, Hügl B, et al. Reduction of atrial fibrillation burden by atrial overdrive pacing: Experience with an improved algorithm to reduce early recurrences of atrial fibrillation. Europace 2009; 11:62–69.
25. Padeletti L, Pürerfellner H, Adler SW, Waller TJ, Harvey M, Horvitz L, Holbrook R, et al. Combined efficacy of atrial septal lead placement and atrial pacing algorithms for prevention of paroxysmal atrial tachyarrhythmia. J Cardiovasc Electrophysiol 2003; 14:1189–1195.
26. Pokushalov E, Romanov A, Corbucci G, Bairamova S, Losik D, Turov A, Shirokova N, et al. Does atrial fibrillation burden measured by continuous monitoring during the blanking period predict the response to ablation at 12-month follow-up? Heart Rhythm 2012; 9:1375–1379.
27. Strickberger SA, Ip J, Saksena S, Curry K, Bahnson TD, Ziegler PD. Relationship between atrial tachyarrhythmias and symptoms. Heart Rhythm 2005; 2:125–131.
28. Quirino G, Giammaria M, Corbucci G, Pistelli P, Turri E, Mazza A, Perucca A, et al. Diagnosis of paroxysmal atrial fibrillation in patients with implanted pacemakers: Relationship to symptoms and other variables. Pacing Clin Electrophysiol 2009; 32:91–98.
29. Page RL, Wilkinson WE, Clair WK, McCarthy EA, Pritchett EL. Asymptomatic arrhythmias in patients with symptomatic paroxysmal atrial fibrillation and paroxysmal supraventricular tachycardia. Circulation 1994; 89:224–227.
30. Israel CW, Grönefeld G, Ehrlich JR, Li Y-G, Hohnloser SH. Long-term risk of recurrent atrial fibrillation as documented by an implantable monitoring device: Implications for optimal patient care. J Am Coll Cardiol 2004; 43:47–52.
31. Israel CW, Neubauer H, Olbrich H-G, Hartung W, Treusch S, Hohnloser SH; BEATS Study Investigators. Incidence of atrial tachyarrhythmias in pacemaker patients: Results from the Balanced Evaluation of Atrial Tachyarrhythmias in Stimulated patients (BEATS) study. Pacing Clin Electrophysiol 2006; 29:582–588.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

A. Detailed patient demographics.
B. Technical details of the computer simulation procedure.
C. Details on the calculation of the atrial fibrillation density (temporal aggregation pattern).
D. Details on the calculation of the relative mean absolute deviation (RAD).
E. Details on the statistical modeling.
F. Table 1: Response Surface Models.
G. References.