AF Detection Performance Comparison: BIOMONITOR vs. LINQ

BIOTRONIK
excellence for life

Table of Contents

1 OVERVIEW

The purpose of this report is to evaluate Atrial Fibrillation (AF) detection algorithm performance in two leading implantable cardiac monitors (ICMs): BIOTRONIK BIOMONITOR and Medtronic Reveal LINQ in an identical patient population.

2 BACKGROUND

AF is the most common cardiac arrhythmia diagnosed and treated in the US and EU [1-2]. In patients with suspected or known AF, it is important to accurately identify the occurrence of AF. The treatment of patients' symptoms as well as the prevention of stroke and heart failure is dependent on accurate detection and characterization of AF. Long-term monitoring using ICMs has been widely used to identify the occurrence and duration of AF, and to manage patients who are already diagnosed with AF, or who have been indicated for AF ablation procedures [3]. This long-term monitoring of AF provides an accurate characterization of the patient's AF, which could significantly impact clinical treatment decisions [3-4]. However, a clear comparison between the leading ICMs has not been conducted on the performance of the AF algorithms.

The performance of the AF detection algorithm is critical to identifying the presence of any AF in a particular subject, and then quantifying the episode durations and total burden of AF during the day. BIOMONITOR provides up to six AF snapshots (ECG strips) of 60 second duration to the physician to make the process of confirming the presence of AF more effective. The AF detection algorithm remains unchanged between BioMonitor 2 and BIOMONITOR III. The performance of AF detection algorithm remains agnostic to the BIOMONITOR device under test as the algorithms are identical.

BIOMONITOR and LINQ utilize proprietary algorithms to detect AF and publications on algorithm performance have been dominated by single arm studies of either one or the other device [5]. Commonly used and accepted performance criteria include Sensitivity (SNS), the ability to detect clinically significant episodes and Positive Predictive Value (PPV), which is an indicator of the clinical burden for review of transmitted episodes. PPV, in particularly, is very dependent on the balance of true and false episodes and is sensitive to a low AF prevalence in the study population. However, inclusion and selection criteria, particularly with regard to prevalence of AF, are quite different among the available studies [5]. The incidence of true positive episodes varies widely within a given patient population or enrollment group, and this can have a significant impact on performance metrics.

Understanding how well AF detection algorithms perform on real clinical data in a known and well characterized patient population is difficult without large randomized trials, and blinding is made challenging by the fact that the implants and their data are visibly different to reviewers and adjudicators. Therefore, even clinical trials do not necessarily provide an accurate side-by-side comparison.

3 METHODS

The objective of this analysis was to compare AF detection performance of BIOMONITOR to that of LINQ in an identical patient population with identical input data in order to quantify true clinical performance of the two algorithms. Therefore, the main objective of this study was to assess the AF burden, duration detection, and episode sensitivity and PPV between the two devices.

3.1 Study Design

This retrospective analysis was conducted using the Holter ECG data collected as part of the BIOTRONIK "AF Detect" study. The study was a prospective single center single-arm study, conducted at the GVM Care & Research Center of the Maria Cecilia Hospital in Cotignola, Italy from November, 2013 to March, 2014. The study was prospectively designed to enroll 66 patients [6]. A subset of the first 50 enrolled patients was used as part of this retrospective analysis, consistent with the dataset used in previous analyses of BIOMONITOR. The patients enrolled in the study: (i) had documented AF episodes or symptoms attributable to AF, (ii) were scheduled for catheter ablation, or (iii) had undergone catheter ablation but were still experiencing AF-related symptoms. All patients met at least one of the above mentioned criteria. Those presenting with long-standing persistent or permanent AF were excluded. The clinical study was conducted in accordance with the Declaration of Helsinki and the international standard for clinical investigation of medical devices in human subjects, ISO 14155. An ethics committee approved the study, and all patients gave written informed consent.

3.2 Study Population

The baseline characteristics of the study population are as follows, the mean age of the cohort was 60.08 ± 9.30 years. There were 9 female and 41 male patients enrolled in the study. The BMI for the enrolled cohort was in the range 21.36 – 38.7, with the median at 26.62 kg/m^2. There was a history of paroxysmal AF in 70% of the enrolled cohort and the rest of the cohort presented with a history of persistent AF. Patients were classified by the severity of their symptoms into the commonly used, New York Heart Association (NYHA) function classification. The cohort level classification is presented in Figure 1.

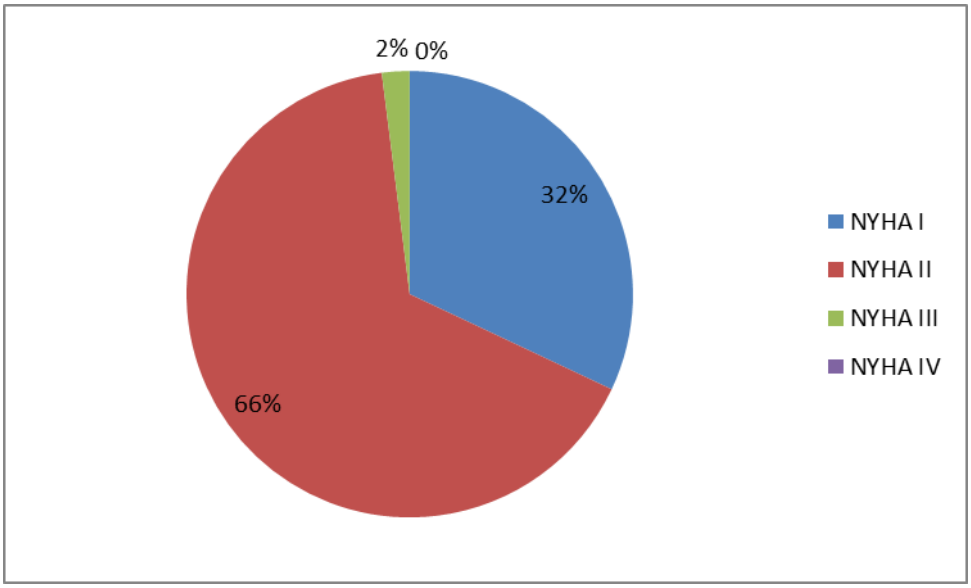

Figure 1: Distribution of study cohort into the New York Heart Association (NYHA) functional classification

3.3 Study Data Collection

After enrollment, each BioMonitor-implanted participant was equipped with a LifeCard Holter (product of Spacelabs Healthcare and FDA-approved compact Holter Ambulatory ECG Recorder) and subsequently participated in a two-day (48 hour), out-of-clinic recording. Holter ECG was used to establish an independent assessment of AF of any episode longer than two minutes, and represented the gold standard for AF detection. The Holter ECGs were manually annotated for AF episode onset and termination. The 48-h Holter ECG signal was evaluated by at least two experts blinded to ICM detections and patient-related information, and included physician review and final adjudication. Segments with non-interpretable Holter ECG due to noise or artifacts were excluded.

3.4 Data Replay Methodology

BIOMONITOR and LINQ were each evaluated against the expert annotations of AF episodes from Holter ECG recordings (~48 hours). True AF episodes that were 6 minutes or longer were selected for data replay and were associated with 28 unique patients with different levels of AF in the study. The database contained a total of 1,209 hours of data of which 709 hours were annotated as AF. The patient averaged AF burden was 25.3 hours, with a median duration of 27.5 hours and with AF duration ranging between 16.5 minutes to 46.3 hours. The individual patient level AF duration is depicted in Figure 2.

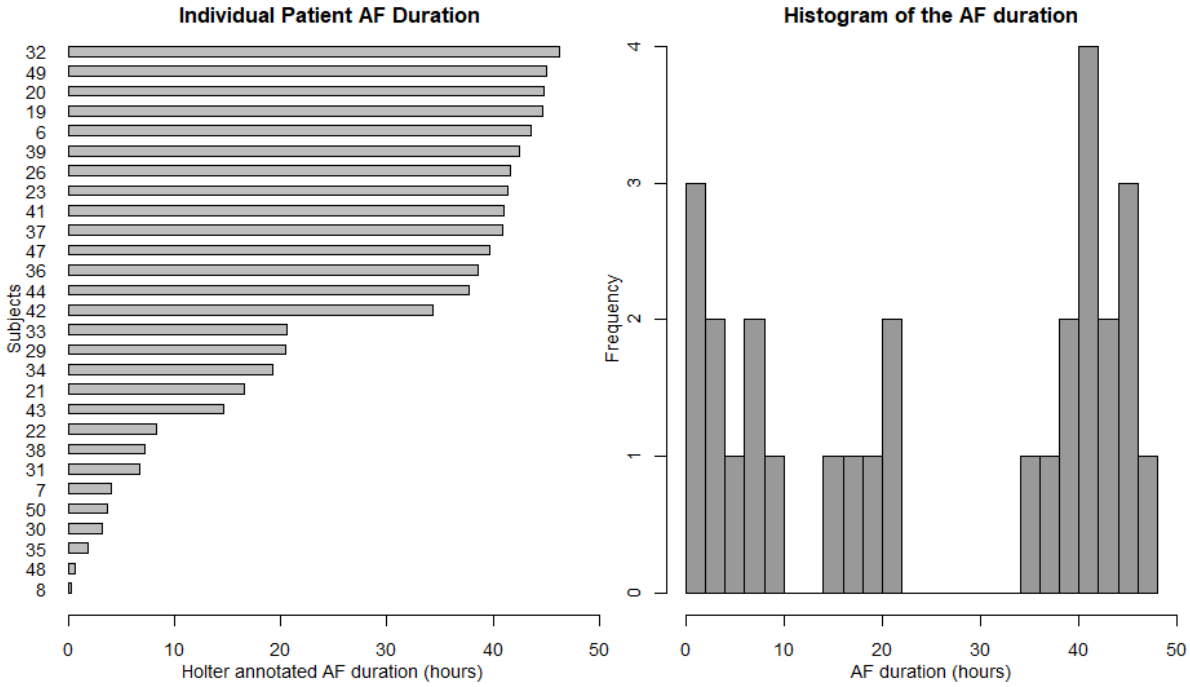

Figure 2. Characteristics of the annotated AF Holter data used in this analysis.

The Holter data was replayed directly into the sensing electrodes of the BIOMONITOR and LINQ devices simultaneously. The BIOMONITOR device used for the current experimental setup was BioMonitor 2; however, the AF algorithm remains unchanged between BioMonitor 2 and BIOMONITOR III. The results are applicable to both devices.

The peak to peak voltage was set to 0.6mV for each replay file for both ICMs. This was determined to be clinically realistic signal amplitude based on 1-month QRS amplitudes reported in the literature for LINQ implants [7]. It was recognized that LINQ reports peak to peak signal amplitudes, while BIOMONITOR reports peak QRS complex amplitudes. For BIOMONITOR, the average reported peak amplitude is around 0.7mV. This smaller input amplitude is a recognized potential disadvantage of the replay methodology for the BIOMONITOR device performance, but would only negatively impact the BIOMONITOR results. The experimental setup is shown in Figure 3.

The AF detection parameters were set comparably between the two devices. Based on the available literature and suggested AF detection settings in [8-9], the AF detection for LINQ were set to balanced sensitivity (P-SENSE enabled) with aggressive ectopy rejection. To maintain parity, the AF sensitivity setting in BIOMONITOR was set to medium with aggressive bigeminy rejection. Episode length was set in both devices to 6 minutes, based on available recommendations [4].

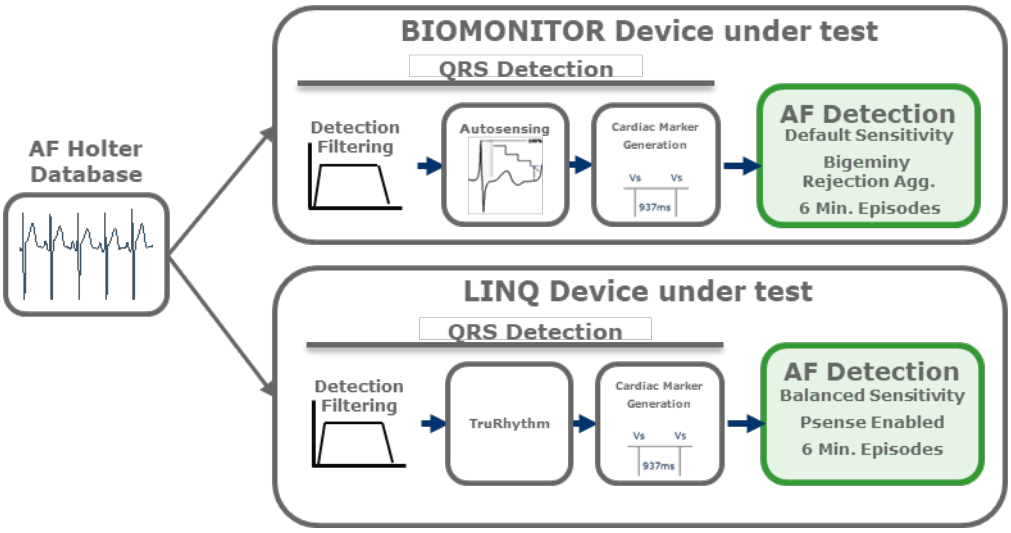

Figure 3. Experimental setup to replay the AF Holter data into the respective ICMs.

3.5 AF Episode Classification

Physician annotations of the AF episodes were used to evaluate the performance of the AF detection algorithms. The ICMs (BIOMONITOR and LINQ) and Holter AF episodes were directly compared to one another on an episode-by-episode basis, by which each episode was classified as true positive (TP), false negative (FN), or false positive (FP), as shown in Figure 4. A Holter AF episode without a coincident ICM AF episode was classified as a FN, while an ICM AF episode occurring without a coincident Holter AF episode was classified as a FP. True positive (TP) episodes were further categorized into one of two types: a Holter AF episode coincident with at least one ICM AF episode was classified as a $TP_{Holter,}$ while an ICM AF episode coincident with at least one Holter AF episode was classified as a TP_{ICM}. These categories $TP_{Holter,}$ and TP_{ICM} are further elucidated by the Figure 4.

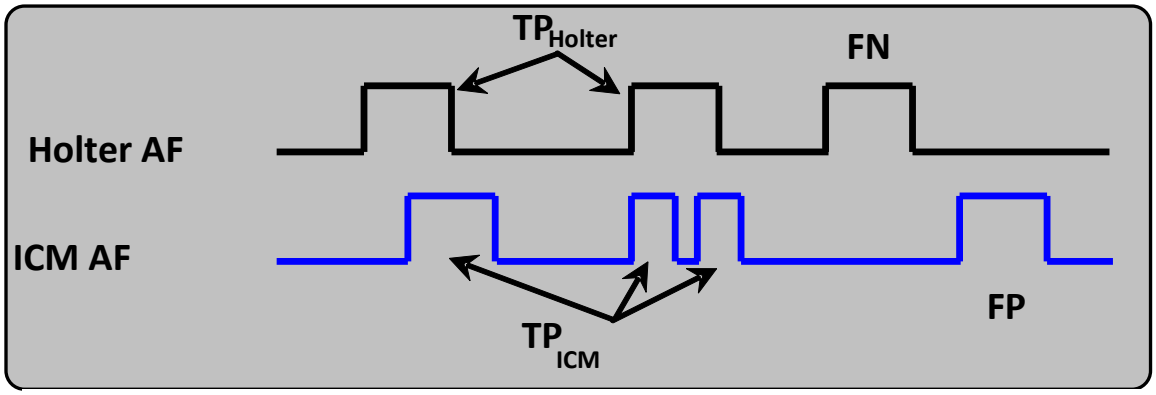

Figure 4. Classification of ICM (BIOMONITOR and LINQ) and Holter AF events

3.6 Calculation of AF Detection Performance

AF episode classifications (TP, FN, FP) for each participant were used to quantify the performance of BIOMONITOR and LINQ AF detection. The diagnostic performance was evaluated based on the patient-based metrics, namely sensitivity (the percentage of Holter AF episodes that were correctly identified by ICM) and the review burden associated with the detection performance was evaluated using PPV (the percentage of ICM AF episodes that correctly identified a Holter AF episode). Sensitivity and PPV were calculated as follows:

Sensitivity (%) = $100 \times TP_{Holter} / (TP_{Holter} + FN)$

PPV (%) = $100 \times TP_{ICM} / (TP_{ICM} + FP)$

Mean values were calculated among all participants with true AF episodes.

The AF duration was calculated as the sum total of all AF episode durations detected by each ICM. The AF burden determined by the ICM was compared with the reference AF burden by calculating the Pearson correlation coefficient between the paired measurements.

4 RESULTS

Each ICM correctly identified 24 of the 28 patients with true AF episodes. Table 1 shows the sensitivity and PPV performance against the 48-hour Holter data replayed into LINQ and BIOMONITOR devices using the AF detection parameter configurations described previously. TP, FP, and FN AF detection counts are displayed for all 28 patients with true AF episodes adjudicated from the Holter.

	LINQ						BIOMONITOR					
Patient	**TP_{Holter}**	**TP_{ICM}**	**FP**	**FN**	**Sensitivity (%)**	**PPV (%)**	**TP_{Holter}**	**TP_{ICM}**	**FP**	**FN**	**Sensitivity (%)**	**PPV (%)**
6	1	1	0	0	100.0	100.0	1	3	0	0	100.0	100.0
7	1	2	0	0	100.0	100.0	1	1	0	0	100.0	100.0
8	1	1	0	0	100.0	100.0	1	1	0	0	100.0	100.0
19	0	0	0	1	0.0	--	1	4	0	0	100.0	100.0
20	1	3	0	0	100.0	100.0	1	1	0	0	100.0	100.0
21	4	8	0	2	66.7	100.0	5	7	1	1	83.3	87.5
22	1	2	0	0	100.0	100.0	1	1	0	0	100.0	100.0
23	1	1	0	0	100.0	100.0	1	1	0	0	100.0	100.0
26	1	1	0	0	100.0	100.0	1	1	0	0	100.0	100.0
29	3	18	0	0	100.0	100.0	3	17	3	0	100.0	85.0
30	0	0	0	6	0.0	--	0	0	0	6	0.0	--
31	0	0	0	7	0.0	--	0	0	0	7	0.0	--
32	1	1	0	0	100.0	100.0	1	1	0	0	100.0	100.0
33	1	27	0	0	100.0	100.0	1	1	0	0	100.0	100.0

AF Detection Performance Comparison: BIOMONITOR vs. LINQ –19-0398-0A

34	5	16	0	10	33.3	100.0	6	12	0	9	40.0	100.0
35	4	4	0	1	80.0	100.0	3	3	0	2	60.0	100.0
36	1	31	0	0	100.0	100.0	1	1	0	0	100.0	100.0
37	1	19	0	0	100.0	100.0	1	6	0	0	100.0	100.0
38	10	12	1	0	100.0	92.3	10	12	0	0	100.0	100.0
39	1	2	0	0	100.0	100.0	1	3	0	0	100.0	100.0
41	1	3	0	0	100.0	100.0	1	2	0	0	100.0	100.0
42	4	22	0	0	100.0	100.0	3	21	1	1	75.0	95.5
43	5	8	0	11	31.3	100.0	4	5	0	12	25.0	100.0
44	1	1	0	0	100.0	100.0	1	1	0	0	100.0	100.0
47	1	5	0	0	100.0	100.0	1	1	0	0	100.0	100.0
48	2	3	0	0	100.0	100.0	0	0	0	2	0.0	--
49	1	8	0	0	100.0	100.0	1	1	0	0	100.0	100.0
50	0	0	0	1	0.0	--	0	0	0	1	0.0	--
Mean					**78.97**	**99.68**					**77.98**	**98.66**

Table 1. AF episode classification counts with sensitivity and PPV for each patient exhibiting true AF.

Table 2 shows the cumulative duration of true AF episodes detected by each device during replay of the Holter data. The Holter durations were obtained from the expert annotations of the Holter recordings. The LINQ and BIOMONITOR durations are the total duration of true positive AF episodes detected by each device during replay. LINQ detected a total of 531.3 hours (74.9%), while BioMonitor detected 561.9 hours (79.2%), compared to a Holter total AF duration of 709.3 hours.

Patient	Holter (hours)	LINQ (hours)	BIOMONITOR (hours)
6	43.53	43.63	42.52
7	4.01	3.97	4.02
8	0.28	0.27	0.21
19	44.64	0.00	3.20
20	44.84	44.73	44.86
21	16.54	14.47	18.01
22	8.26	8.20	8.20
23	41.36	41.27	41.12
26	41.66	41.67	41.55
29	20.41	8.60	16.34
30	3.17	0.00	0.00
31	6.73	0.00	0.00
32	46.25	46.20	46.22
33	20.62	11.57	20.62
34	19.27	3.83	6.32
35	1.85	1.50	1.41
36	38.55	21.83	38.52

37	40.93	4.33	3.46
38	7.24	7.03	7.25
39	42.44	42.40	36.48
41	40.98	40.97	41.03
42	34.38	19.10	15.87
43	14.64	3.13	2.37
44	37.68	37.67	37.67
47	39.64	39.47	39.65
48	0.67	0.67	0.00
49	45.05	44.80	44.99
50	3.68	0.00	0.00
Total Duration	**709.32**	**531.30**	**561.87**

Table 2. Cumulative duration of true AF episodes, in hours.

Figure 5 illustrates device-detected AF duration versus LifeCard (Holter) AF Duration for true positive LINQ and BIOMONITOR episodes. The identity line (device-detected duration equals Holter duration) is plotted for reference. Pearson correlation coefficients for device-detected AF duration compared to Holter duration are r=0.809 for LINQ and r=0.821 for BIOMONITOR.

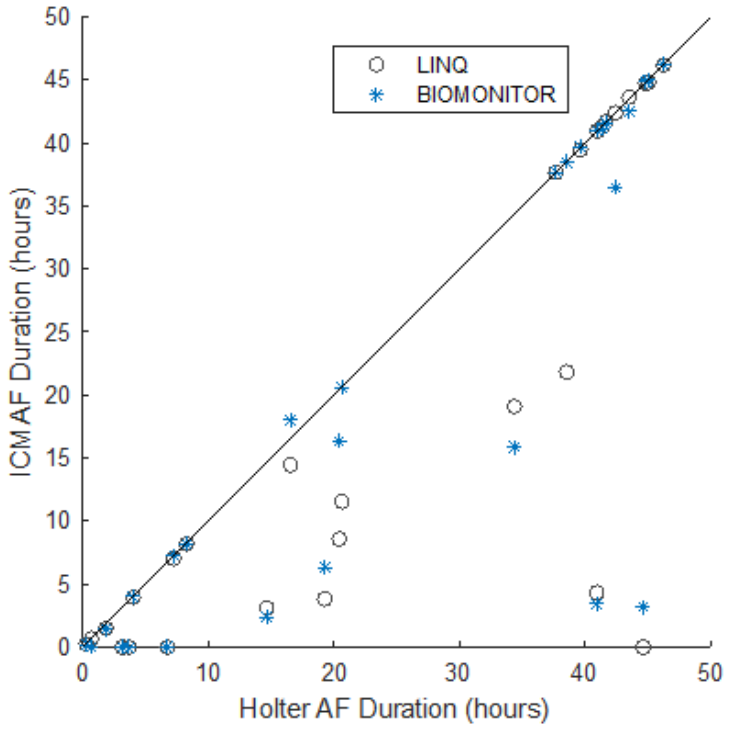

Figure 5. Holter versus AF detected AF duration

5 CONCLUSION

In this first head-to-head comparison of AF detection algorithms, BIOMONITOR and LINQ devices performed with clinical equivalence. Patient-averaged episode sensitivity for BIOMONITOR and LINQ were 78.0% and 79.0%, respectively. Patient-averaged PPV was also within 1% with a 98.7% and 99.7% result for BIOMONITOR and LINQ, respectively. Further, the total duration of classified true AF rhythm compared to total Holter duration was nearly equivalent with BIOMONITOR classifying 79.2% of AF correctly and LINQ classifying 74.9% of AF correctly. In the analysis of the AF burden between the two devices, the evaluation of the AF burden for BIOMONITOR was slightly better than LINQ performance by 4.3%.

This analysis demonstrates clearly that when the two devices analyze the same clinical data, with a single adjudicated data set, performance between the devices is remarkably consistent at a technical level and completely equivalent at the level of the clinical user. The differences in the performance observed in this analysis are put into perspective by the vast range of clinical performance reported in the peer-reviewed literature. LINQ studies have recently been published with patient-averaged episode sensitivity of 97.2% and PPV value of 90.4% [10]. A separate report found patient-averaged episode sensitivity ranging from 86.6% to 90.0% based on patient population [11] and PPV ranging from 78.2% to 89.9% in the same population. Other published data report PPV of 81% among patients with known AF and 39% among patients with cryptogenic stroke [12]. Even within a single device, clinical performance varies considerably based on the distribution of AF prevalence, episode durations, and properties of the AF rhythm within the sampled population.

Although QRS sensing was controlled here by the replay methodology and the same data was input into the two devices, clinically it has been reported that the longer vector of the BIOMONITOR results in higher QRS signal amplitude. This is expected to confer additional advantages in terms of sensing quality (signal to noise ratio) and interpretable AF ECGs in BIOMONITOR when implanted and subject to noise sources that were controlled and thus eliminated here. Based on the preceding comparison and reporting of the BIOMONITOR and LINQ algorithms and in light of normal clinical variations in device performance, the BIOMONITOR and LINQ AF performance is considered equivalent in normal clinical use in a real world AF patient population.

6 CITATIONS

1. Naccarelli, Gerald V., Helen Varker, Jay Lin, and Kathy L. Schulman. 2009. “Increasing Prevalence of Atrial Fibrillation and Flutter in the United States.” The American Journal of Cardiology 104 (11): 1534–39.
2. European Heart Rhythm Association (EHRA), Developed With The Special Contribution of the, Endorsed by The European Association for Cardio-Thoracic Surgery (EACTS), Authors/task Force Members, A. John Camm, Paulus Kirchhof, Gregory Y. H. Lip, Ulrich Schotten, et al. 2010. “Guidelines for the Management of Atrial Fibrillation: The Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC).” European Heart Journal 31 (19): 2369–2429.
3. Lee, Randall, and Suneet Mittal. 2018. “Utility and Limitations of Long-Term Monitoring of Atrial Fibrillation Using an Implantable Loop Recorder.” *Heart Rhythm: The Official Journal of the Heart Rhythm Society* 15 (2): 287–95.
4. European Heart Rhythm Association, European Association for Cardio-Thoracic Surgery, A. John Camm, Paulus Kirchhof, Gregory Y. H. Lip, Ulrich Schotten, Irene Savelieva, et al. 2010. “Guidelines for the Management of Atrial Fibrillation: The Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC).” *European Heart Journal* 31 (19): 2369–2429.
5. Ciconte, Giuseppe, Daniele Giacopelli, and Carlo Pappone. 2017. “The Role of Implantable Cardiac Monitors in Atrial Fibrillation Management.” *Journal of Atrial Fibrillation* 10 (2): 1590.
6. Ciconte, Giuseppe, Massimo Saviano, Luigi Giannelli, Zarko Calovic, Mario Baldi, Cristiano Ciaccio, Amarild Cuko, et al. 2017. “Atrial Fibrillation Detection Using a Novel Three-Vector Cardiac Implantable Monitor: The Atrial Fibrillation Detect Study.” *Europace: European Pacing, Arrhythmias, and Cardiac Electrophysiology: Journal of the Working Groups on Cardiac Pacing, Arrhythmias, and Cardiac Cellular Electrophysiology of the European Society of Cardiology* 19 (7): 1101–8.
7. Pürerfellner, Helmut, Prashanthan Sanders, Evgeny Pokushalov, Marco Di Bacco, Tracy Bergemann, Lukas R. C. Dekker, and Reveal LINQ Usability Study Investigators. 2015. “Miniaturized Reveal LINQ Insertable Cardiac Monitoring System: First-in-Human Experience.” *Heart Rhythm: The Official Journal of the Heart Rhythm Society* 12 (6): 1113–19.
8. Pürerfellner, Helmut, Prashanthan Sanders, Shantanu Sarkar, Erin Reisfeld, Jerry Reiland, Jodi Koehler, Evgeny Pokushalov, Luboš Urban, and Lukas R. C. Dekker. 2018. “Adapting Detection Sensitivity Based on Evidence of Irregular Sinus Arrhythmia to Improve Atrial Fibrillation Detection in Insertable Cardiac Monitors.” *Europace: European Pacing, Arrhythmias, and Cardiac Electrophysiology: Journal of the Working Groups on Cardiac Pacing, Arrhythmias, and Cardiac Cellular Electrophysiology of the European Society of Cardiology* 20 (FI_3): f321–28.
9. Wechselberger, Simon, Mads Kronborg, Yan Huo, Judith Piorkowski, Sebastian Neudeck, Ellen Päßler, Ali El-Armouche, et al. 2018. “Continuous Monitoring after Atrial Fibrillation Ablation: The LINQ AF Study.” *Europace: European Pacing, Arrhythmias, and Cardiac Electrophysiology: Journal of the Working Groups on Cardiac Pacing, Arrhythmias, and Cardiac Cellular Electrophysiology of the European Society of Cardiology* 20 (FI_3): f312–20.
10. Sanders, Prashanthan, Helmut Pürerfellner, Evgeny Pokushalov, Shantanu Sarkar, Marco Di Bacco, Bärbel Maus, Lukas R. C. Dekker, and Reveal LINQ Usability Investigators. 2016. “Performance of a New Atrial Fibrillation Detection Algorithm in a Miniaturized Insertable Cardiac Monitor: Results from the Reveal LINQ Usability Study.” *Heart Rhythm: The Official Journal of the Heart Rhythm Society* 13 (7): 1425–30.
11. Pürerfellner, Helmut, Evgeny Pokushalov, Shantanu Sarkar, Jodi Koehler, Ren Zhou, Lubos Urban, and Gerhard Hindricks. 2014. “P-Wave Evidence as a Method for Improving Algorithm to Detect Atrial Fibrillation in Insertable Cardiac Monitors.” *Heart Rhythm: The Official Journal of the Heart Rhythm Society* 11 (9): 1575–83.
12. Mittal, Suneet, John Rogers, Shantanu Sarkar, Jodi Koehler, Eduardo N. Warman, Todd T. Tomson, and Rod S. Passman. 2016. “Real-World Performance of an Enhanced Atrial Fibrillation Detection Algorithm in an Insertable Cardiac Monitor.” *Heart Rhythm: The Official Journal of the Heart Rhythm Society* 13 (8): 1624–30.