EUROPEAN MEDICINES AGENCY

SCIENCE MEDICINES HEALTH

5 February 2016
EMA/CHMP/87883/2016 Corr.[1]
Committee for Medicinal Products for Human Use (CHMP)

Assessment report

Xarelto

International non-proprietary name: rivaroxaban

Procedure No. EMEA/H/C/000944/LEG-037

Note

Variation assessment report as adopted by the CHMP with all information of a commercially confidential nature deleted.

---

[1] 10 February 2016

Table of contents

1. Introduction

This report covers a recently recognized post-authorisation signal regarding the device used for INR testing in the Rocket AF and J Rocket studies. The Rocket AF study was the pivotal phase III study supporting the indication for thromboembolic prophylaxis in atrial fibrillation.

On Sept 23, 2015 the EMA received a letter from the MAH in relation to issues with the INR testing methodology applied in these two trials.

The manufacturer of the point of care INR device (Alere Inc.TM) had identified that the device in certain clinical conditions gave inappropriately low INR values. The differences were in some instances reported to be clinically relevant leading to increases of anti-vitamin K agents with resulting increased bleeding tendency. In principle, depending of the size of the problem, this could lead to an increased bleeding rate in the control arm of the trials which would potentially hamper the interpretations of the differences between treatment arms.

Upon receipt of the information a specific regulatory procedure (LEG) was initiated dedicated to the device in question had been used in the above mentioned trials. The strategy to analyse the data from the Rocket AF trial where discussed and questions raised to the MAH in October as preliminary clarification of the issue. (see below).

Request 1:

Clarification on the reasons for the delay in informing the EMA, considering that more than 9 months elapsed since the US alert and recall. Clarification on whether alerts or recall have taken place any other countries, including EU. In addition, exact details of the grounds and extent of the medical device alert and recalls should be provided.

Request 2:

The MAH's view on the impact on the results reported in the ROCKET trials. The Company proposals for sensitivity analyses to be submitted by 15 October 2015 are supported. Further analyses might be requested upon assessment of the data.

Request 3:

Clarification whether the recalled device was used in other trials performed with Xarelto or any other development programmes, and submitted or intended for submission.

Request 4:

Information whether other Regulatory authorities have been informed and subsequent actions were taken in any territory.

During the review the MAH was asked by CHMP to provide additional sensitivity analyses which are further discussed in the scientific part of this report.

1.1. Steps taken for the assessment of data

Timetable	Planned dates
Start of procedure:	15 Oct 2015
Submission of the responses by MAH	15 Oct 2015
CXMP Rapporteur Assessment Report	29 Oct 2015
CHMP discussion	November 2015
CHMP discussion	December 2015
Submission of answers by MAH	7, 18, 22 December 2015 18 January 2016
Rapporteur Revised Assessment Report	8 , 19 and 22 January 2016
CHMP discussion	January 2016

1.2. Background information

On Dec 10, 2014 Alere Inc initiated a voluntary correction to inform users of the device that the system may provide INR results that are clinically significantly lower than results obtained using a reference INR system (laboratory method). A field safety notice was distributed to the National Competent Authorities in the affected European countries.

The message stated that the issue can arise if the patient has certain clinical conditions or if the instructions in the labelling for performing the test are not followed.

It was recommended by the manufacturer that the system should not be used in patients with any of the following conditions:

- Anaemia of any type with haematocrit less than 30%

- Any conditions associated with fibrinogen levels including:

    – Acute inflammatory conditions (examples may include acute viral or bacterial infections such as pneumonia or influenza)

    – Chronic inflammatory conditions (examples may include rheumatoid arthritis, Crohns disease, ulcerative colitis, infectious liver disease such as hepatitis, or inflammatory kidney disease such as diabetic nephropathy and glomerulonephritis)

    – Severe infection (e.g.sepsis)

    – Chronically elevated fibrinogen levels for any reasons

    – Hospitalized or advanced stage cancer or end stage renal disease patients requiring haemodialysis

- Any bleeding or unusual bruising, clinically observed or reported by the patient

Patients with any of the conditions listed above were recommended to immediately be transitioned to a laboratory INR method for monitoring of their warfarin therapy.

2. Scientific assessment of the information provided by the MAH

2.1. Delay in informing the EU regulatory Authorities

Clarification on the reasons for the delay in informing the EMA, considering that more than 9 months elapsed since the US alert and recall.

Summary of MAH response

The ROCKET AF trial

In order to maintain the blinding of investigators to the study medication in the Rocket AF trial the necessary INR control for warfarin was performed with an INR point of care device, which was provided to the sites by Janssen Research & Development LLC (JRD), Bayer's co-development partner for rivaroxaban. For this purpose, JRD purchased over 3,000 INRatio devices from HemoSense, Inc., (San Jose, CA) and Inverness Medical Innovations, Inc. (Waltham, MA) between 2006 and 2009. The devices were specifically dedicated for the use in the ROCKET AF trial (for details please refer to Response to Request 2).

The Alere INR ratio device recall

In August 2007, HemosSense, the original provider of the device used in the ROCKET AF trial, was purchased by Inverness Medical Innovations, Inc., an US- based (Waltham, MA) biotechnology company. The company changed its name to Alere in the year 2010 and continued to market the INRatio devices under the name Alere INRatio and INRatio2 PT/INR Monitor System.

On December 5, 2014 a class 1 recall was issued by FDA for the Alere INRatio and INRatio 2 PT/INR monitor systems and test strips, mandating an urgent medical device correction letter (no withdrawal of devices from the market). According to information from Alere to Bayer and JRD a field safety action notice was filed to countries outside the US.

The urgent medical device correction letter was sent by Alere Inc. to users of its INRratio and INR Ratio 2 PT/INR monitor systems and test strips that were manufactured between April 2008 and December 2014, relating to all lots of part numbers: 0100071 and 0100139.

The part numbers of the devices purchased from HemoSense Inc. for use in the ROCKET AF trial (part 200361) are different than those listed in the recall. Bayer's development partner JRD was not directly informed by the successor company Alere Inc. at any time in the context of the recall. Alere only reached out to clients that purchased the devices after 2009.

On September 9, 2015, JRD became aware through a 3rd party that the recall notice of the Alere INRatio device might be applicable to the HemoSense INRatio devices used in the ROCKET AF trial program as well. Only through an immediate active request to Alere by JRD and a subsequent research by Alere in their legacy files, a connection between the lots cited in the official public recall notice and the unique identifier used on the ROCKET AF trial program devices was revealed. Subsequently Alere confirmed in a letter dated September 24, 2015 that the Urgent Medical Device Correction notification for the Alere TM

INRatio®PT/INR Monitor System did contain information which is applicable to the PT/INR Monitor System PN 200361 used in the studies conducted by JRD. They further clarified why the devices used in the ROCKET AF trial were not listed in their original notification: "Alere generated the affected parts list attached to the notification and the customer consignee list based on a distribution database which contained distribution data back to 2009. The 200361 part began distribution in 2006 and was not included in this distribution database. Given the INRatio Test Strip dating is 15 months, this timeframe ensured all current customers were notified."

Bayer and JRD informed the respective Health Authorities per local regulations and practices in their respective territories (JRD for USA, Bayer outside USA) in parallel, as soon as the applicability of the device recall was confirmed by Alere.

In summary, the sequence of events as described above including the corresponding background documents shows that neither Bayer nor its co-development partner JRD had become aware of the potential applicability of the correction/recall (recall) to the Hemonsense Devices used in ROCKET AF before Sep 9, 2015. A connection between the recall and the devices purchased for the Rocket AF trial could not be made by Bayer or JRD based on publicly available information.

- The devices had been purchased from the acquired company HemoSense and the names in the recall notice did not match with the name of the devices used in the Rocket AF trial.

- The timeframe given in the recall notice of December 2014 began after the main purchase order for the trial.

- The lot/part numbers listed in the recall did not match with the identifiers of the devices used in the ROCKET AF trial.

- JRD was not directly informed by Alere at any time in the context of the recall. Alere only reached out to those clients that purchased the devices after 2009.

- After becoming aware of the potential applicability of the recall to the ROCKET AF study through a third party on 9 September 2015, JRD actively engaged Alere so that a search in Alere's legacy files was conducted.

- Alere informed JRD that the Urgent Medical Device Correction notification did contain information which was applicable to the PT/INR Monitor system PN 200361 in a formal letter on 24 September 2015.

Clarification on whether alerts or recall have taken place any other countries, including EU

As noted on the FDA website the devices were being recalled in patients with certain conditions as they "may provide an INR result that is lower than expected result obtained using a laboratory INR method". The same advice was provided by other European Agencies.

Bayer is neither a manufacturer nor a distributor of the INRatio and INRatio 2 PT/INR Monitor System and the device was not chosen to monitor INR using in any other development program other than the ROCKET AF trial program. The EMA request for more information on the device has been forwarded to Alere.

On October 08, 2015 Alere sent an email to JRD explaining that "the INRatio Monitor product notice was also distributed in the European countries which received the product. The recall notice is in the form of a Field Safety Notice (FSN)." Furthermore Alere provided the Field Safety Notice (FSN) as well as a list of European countries the field notice was distributed to, as listed below:

Austria, Belgium, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Lithuania, Malta, Netherlands, Norway, Poland, Portugal, Romania, Spain, Switzerland, Turkey, United Kingdom.

In addition, exact details of the grounds and extent of the medical device alert and recalls should be provided.

According to the voluntary urgent correction notice issued by Alere the device should not be used on patients with certain medical conditions (see above).

As further explained in the response to request 2, this correction notice was used by Bayer and JRD as the basis to propose and define the sensitivity analyses performed. In the response to request 2, Bayer addresses the question on how the device was used in the ROCKET AF study and assesses the potential implications of the device use on the reliability of the clinical data supporting the demonstration of the positive risk-benefit profile for Xarelto in the AF indication.

Upon confirmation by Alere on September 24, 2015 on the applicability of the device recall to the ROCKET AF study, JRD followed up with Alere in verbal and written communication in order to obtain more information on the grounds and extent of the recall. The request of EMA on 'the exact details of the grounds and extent of the medical device alert and recalls' was forwarded to Alere on October 5th, 2015. On October 8th, 2015, Alere provided the INRatio Monitor product notice that was distributed in the European countries where the product is marketed. The recall notice is in the form of a Field Safety Notice (FSN). The technical information in the FSN is the same as in the US notice to FDA. This FSN was reported to the European Competent Authorities under FSCA 2014/012/015/081/005. In the same e-mail Alere informed JRD of the following:

"*A Health Hazard Assessment was conducted which concluded that the occurrence rate of a discrepant low result was occasional and the probability of injury as a result of the discrepancy is unlikely but possible. With INR testing, there is the possibility of a severe or life-threatening injury but the overall risk was determined to be Moderate. Alere has communicated to customers that more than 99% of all patient INR results are not expected to be affected by this specific issue.*

We have communicated more detailed information directly to FDA and EU Competent Authorities. Because of the proprietary nature of this information, we would directly provide this information to EMA."

After receipt of the information from Alere, Bayer immediately informed EMA and also provided respective contact details to Alere in order to ensure process support for submissions of confidential information from Alere directly to EMA.

Bayer informed EMA on October 09, 2015 about Alere´s support in this matter and asked for clarification on the procedural aspect of this submission by Alere. These were clarified by EMA on October 12, 2015 with the specific request to inform Alere on the possibility that the assessment report prepared by EMA may contain elements of the information provided by Alere to EMA, before Alere provides information to the EMA. This information was shared with Alere on October 12, 2015 per email and Alere reiterated their support on October 12, 2015.

Further relevant information on the grounds and extent of the medical device alert and recalls will be provided directly from Alere to EMA.

In conclusion, the MAH has clarified the sequence of actions taken by the different actors. Based on the response it can be concluded that the MAHs was not aware of any potential impact of the identified deficiencies of INR device system on the Rocket studies until Sept 9, 2015. The clarifications are accepted by CHMP.

However, it should be emphasized that it is the responsibility of Bayer, as MAH, to ensure that it maintains oversight of the safety profile and is kept informed of major issues, including the use of devices used during trials. In addition, appropriate traceability of data generated by devices used during a study should be ensured by the sponsor and retained even after the study has been finalised.

2.2. Assessment on the scientific data requested by CHMP

Request 2: The MAH's view on the impact on the results reported in the ROCKET trials is requested and the Company proposals for sensitivity analyses submitted by 15 October 2015 are supported. However, further analyses might be requested upon assessment of the data.

Summary of the MAH response:

Background information on ROCKET AF

ROCKET AF was a multicenter, randomized, double-blind, double-dummy, event-driven trial that was conducted at 1178 participating sites in 45 countries. The study was financially supported by Johnson & Johnson R&D (now Janssen R&D), and Bayer. The primary objective of the study was to demonstrate that the efficacy of rivaroxaban is non-inferior to that of dose adjusted warfarin for the prevention of thromboembolic events in subjects with non-valvular atrial fibrillation as measured by the composite of stroke (including primary ischemic, primary haemorrhagic, and unknown type unless otherwise stated) and non-central nervous system (CNS) systemic embolism. The principal safety objective of the study was to demonstrate that rivaroxaban is superior to dose-adjusted warfarin as assessed by the composite of major and non-major clinically relevant bleeding events.

Patients were randomly assigned to receive fixed-dose rivaroxaban (20 mg daily or 15 mg daily in patients with a creatinine clearance of 30 to 49 ml per minute) or adjusted-dose warfarin (target international normalized ratio [INR], 2.0 to 3.0). Patients in each group also received a placebo tablet in order to maintain blinding. Randomization was performed with the use of a central 24-hour, computerized, automated voice-response system. To maintain the blind in this double-blind, double dummy study, warfarin and its matching placebo were dose-adjusted based on either real or sham INR results, respectively. Study sites were provided with 2 each of the point-of-care INR devices (INRatio) that performed the prothrombin time and INR based on the subject's whole blood sample. The aim of using these point-of-care devices for INR measurement in this double-blind double dummy study was 2-fold: 1) to capitalize on their capability to provide encryption of each INR value obtained and thereby improve the shamming and blinding capability for the trial, and 2) to facilitate more rapid and real time turnaround of INR values (as opposed to sending a lab specimen away to a laboratory) and the performance of dose adjustment of warfarin, if needed, as soon as possible.

Summarizing the efficacy results, in the primary analysis (per protocol/on-treatment), the primary end point occurred in 188 patients in the rivaroxaban group (1.7% per year) and in 241 in the warfarin group (2.2% per year) (hazard ratio in the rivaroxaban group, 0.79; 95% confidence interval [CI], 0.66 to 0.96; $P<0.001$ for non inferiority). In the intention-to-treat analysis (Table 1), the primary end point occurred in 269 patients in the rivaroxaban group (2.1% per year) and in 306 patients in the warfarin group (2.4% per year) (hazard ratio, 0.88; 95% CI, 0.74 to 1.03; $P<0.001$ for non-inferiority; $P = 0.12$ for superiority). Other secondary efficacy analyses demonstrated there were consistently fewer events, strokes, non-CNS systemic embolisms, vascular deaths, and MIs observed for the rivaroxaban group compared with warfarin group. Additionally, the treatment effects were consistent across populations and subgroups.

Table 1. Efficacy analyses results of ROCKET AF as summarized in the SPC

Study population	ITT analyses of efficacy in patients with non-valvular atrial fibrillation		
Treatment dosage	Xarelto 20 mg od (15 mg od in patients with moderate renal impairment) Event rate (100 pt-yr)	Warfarin titrated to a target INR of 2.5 (therapeutic range 2.0 to 3.0) Event rate (100 pt-yr)	Hazard ratio (95% CI) p-value, test for superiority
Stroke and non-CNS systemic embolism	269 (2.12)	306 (2.42)	0.88 (0.74 - 1.03) 0.117
Stroke, non-CNS systemic embolism and vascular death	572 (4.51)	609 (4.81)	0.94 (0.84 - 1.05) 0.265
Stroke, non-CNS systemic embolism, vascular death and myocardial infarction	659 (5.24)	709 (5.65)	0.93 (0.83 - 1.03) 0.158
Stroke	253 (1.99)	281 (2.22)	0.90 (0.76 - 1.07) 0.221
Non-CNS systemic embolism	20 (0.16)	27 (0.21)	0.74 (0.42 - 1.32) 0.308
Myocardial infarction	130 (1.02)	142 (1.11)	0.91 (0.72 - 1.16) 0.464

In the safety population (Table 2), the incidence of CEC-adjudicated bleeding events was comparable for the principal safety endpoint of major and non major clinically relevant bleeding (20.74% for rivaroxaban and 20.34% for warfarin) and there was no statistically significant difference between the treatment groups with a hazard ratio of 1.03 (95% CI 0.96 to 1.11, p = 0.442). The bleeding event rates for the principal safety endpoint were comparable between treatment groups (14.91/100 patient-years rivaroxaban, 14.52/100 patient-years warfarin) as were those for the major (3.60/100 patient-years rivaroxaban, 3.45/100 patient-years warfarin) and non-major clinically relevant (11.80/100 patient-years rivaroxaban, 11.37/100 patient-years warfarin) bleeding events separately. Significant reductions in intracranial haemorrhage (0.5% vs. 0.7%, P = 0.02) and fatal bleeding (0.2% vs. 0.5%, P = 0.003) in the rivaroxaban group were observed.

Table 2. Safety analyses results of ROCKET AF as summarized in the SPC

Study population	Patients with non-valvular atrial fibrillation[a]		
Treatment dosage	Xarelto 20 mg once a day (15 mg once a day in patients with moderate renal impairment) Event rate (100 pt-yr)	Warfarin titrated to a target INR of 2.5 (therapeutic range 2.0 to 3.0) Event rate (100 pt-yr)	Hazard ratio (95% CI) p-value
Major and non-major clinically relevant bleeding events	1,475 (14.91)	1,449 (14.52)	1.03 (0.96 - 1.11) 0.442
Major bleeding events	395 (3.60)	386 (3.45)	1.04 (0.90 - 1.20) 0.576
Death due to bleeding*	27 (0.24)	55 (0.48)	0.50 (0.31 - 0.79) 0.003
Critical organ bleeding*	91 (0.82)	133 (1.18)	0.69 (0.53 - 0.91) 0.007
Intracranial haemorrhage*	55 (0.49)	84 (0.74)	0.67 (0.47 - 0.93) 0.019
Haemoglobin drop*	305 (2.77)	254 (2.26)	1.22 (1.03 - 1.44) 0.019
Transfusion of 2 or more units of packed red blood cells or whole blood*	183 (1.65)	149 (1.32)	1.25 (1.01 - 1.55) 0.044
Non-major clinically relevant bleeding events	1,185 (11.80)	1,151 (11.37)	1.04 (0.96 - 1.13) 0.345
All cause mortality	208 (1.87)	250 (2.21)	0.85 (0.70 - 1.02) 0.073

a) Safety population, on treatment
* Nominally significant

Potential pertinence of Alere Correction/Recall to the ROCKET AF trial

The INRatio device was modified by the manufacturer (HemoSense) with a software program to encrypt the INR reading and maintain blinding. The read-out on the screen was a 7-digit code that corresponded to the actual or shammed INR value (there were multiple 7-digit codes assigned to each INR value to prevent inadvertent unblinding by the investigative site). To obtain an INR value for that individual test, once the investigative site performed the INR on a patient they called the IVRS or accessed the IWRS and entered the subject's sudy identification number. After entering the identifying information as well as the last three doses of warfarin or warfarin placebo, the investigator entered the 7-digit code obtained from the INRatio device. If the subject was randomized to warfarin, the INR value reported by the IVRS/IWRS was the actual value. If the subject was randomized to rivaroxaban (with warfarin placebo) the INR value given by the IVRS/IWRS was a sham value that mimicked values obtained as if the subject were taking warfarin.
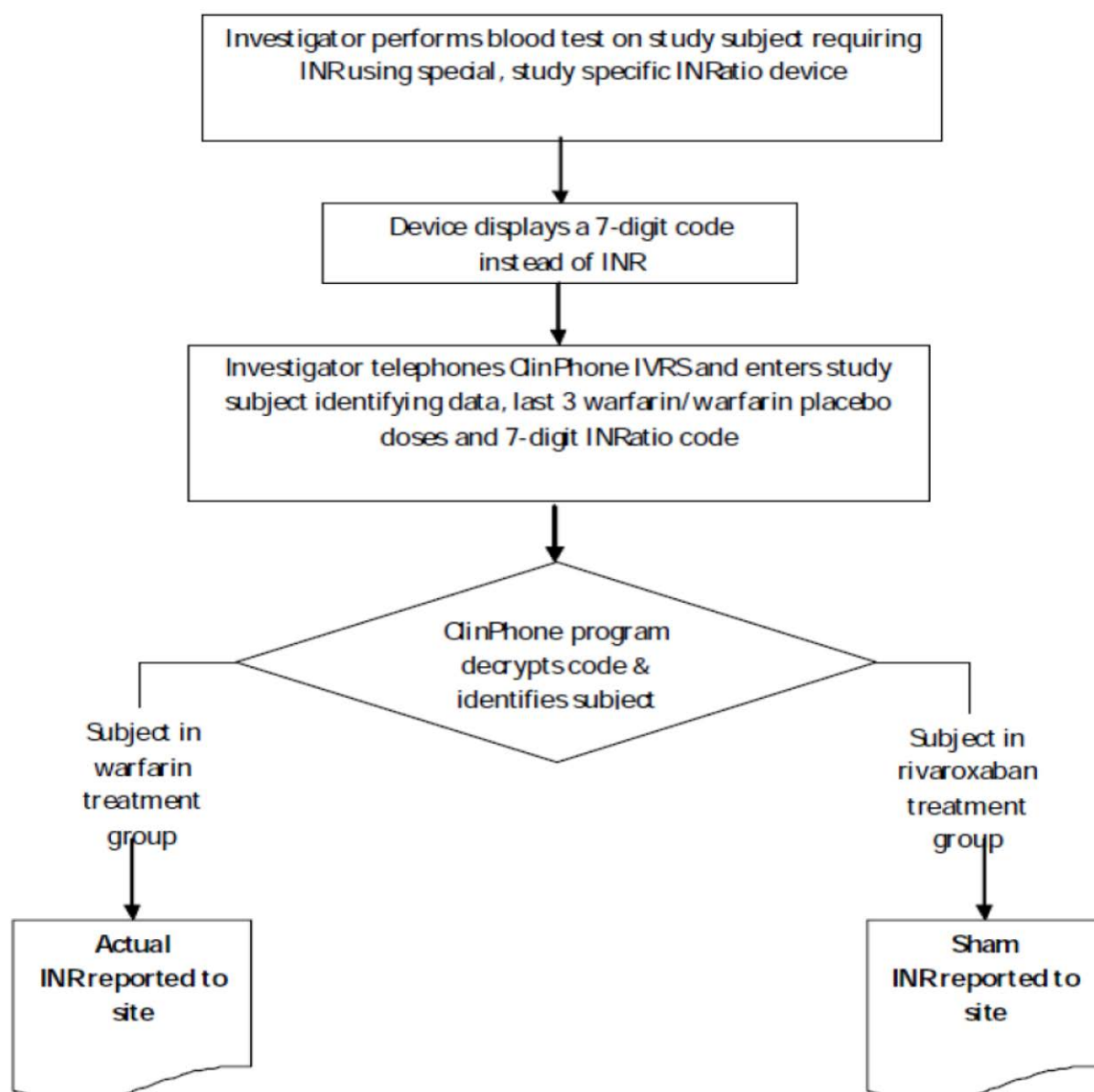
The INR provided by the IVRS was to be used by the investigator to adjust the warfarin/warfarin placebo dose in all subjects from the time of randomization until transition from study drug to an open-label VKA or other appropriate therapy. INR monitoring (using the point-of-care devices provided) was to occur as clinically indicated but at least every 4 weeks. Unblinded INR measurements were not to be performed while on study drug, except in case of a medical emergency. Sites were instructed to measure the INR using only the point-of-care devices provided to ensure consistency of warfarin dosing and to preserve the integrity of the study blind.

The INR shamming program developed for the IVRS was derived from a proprietary algorithm developed from literature review and actual patient data from anticoagulation clinics in Sweden which included patients with AF receiving anticoagulation. Thus, investigators in ROCKET AF adjusted study medication (whether active warfarin or placebo warfarin) accordingly to maintain an INR target of 2.5 (range 2.0 to 3.0, inclusive). During investigator meetings, INR monitoring was reviewed and treating physicians were encouraged to achieve and maintain an INR target of 2.5 (range 2.0 to 3.0) for all subjects. In addition, letters and related site communications were sent periodically to all investigators reminding them of the need to maintain INRs within the target range. However, specific warfarin dosing instructions were not provided to the enrolling sites.

A monitor unblinded to study data was employed to review INR data and ascertain if subjects were frequently out of range. This monitor could consult with a physician unblinded to study data at DCRI to discuss specific cases, if needed. Occasionally and as a result of these surveillance efforts, specific investigators whose patients were found to be persistently below or above the target range received correspondence reminding them of the importance of achieving the INR target. The monitor unblinded to study data was also available to answer questions about individual INR results, in a blinded fashion, from investigators, through local medical monitors. At no time did the monitor unblinded to study data evaluate aggregate INR time in therapeutic range. The IDMC also reviewed the INR data on a regular basis. At no time did the IDMC interact with the sites or the INR monitors unblinded to study data.

The use of the INRatio device in the ROCKET AF trial is outlined in Figure 1.

Figure 1. Flow Diagram of the use of the INRatio Device

```
┌─────────────────────────────────────────────────────┐
│ Investigator performs blood test on study subject    │
│ requiring INR using special, study specific INRatio  │
│ device                                               │
└─────────────────────────────────────────────────────┘
                         │
                         ▼
         ┌──────────────────────────────┐
         │ Device displays a 7-digit    │
         │ code instead of INR          │
         └──────────────────────────────┘
                         │
                         ▼
     ┌─────────────────────────────────────────┐
     │ Investigator telephones ClinPhone IVRS   │
     │ and enters study subject identifying     │
     │ data, last 3 warfarin/warfarin placebo   │
     │ doses and 7-digit INRatio code           │
     └─────────────────────────────────────────┘
                         │
                         ▼
                  ◇ ClinPhone program ◇
                  decrypts code &
                  identifies subject
            ┌────────────┴────────────┐
  Subject in                       Subject in
  warfarin                         rivaroxaban
  treatment                        treatment
  group                            group
     │                                │
     ▼                                ▼
┌───────────────┐              ┌───────────────┐
│ Actual        │              │ Sham          │
│ INR reported  │              │ INR reported  │
│ to site       │              │ to site       │
└───────────────┘              └───────────────┘
```

Details of the Sensitivity Analyses

In the communication with EMA and Co-Rapporteurs on 29 September 2015, Bayer proposed to conduct sensitivity analyses to better understand the potential effect of the recall on the results of the ROCKET AF study. It was proposed that a subset of patients with conditions listed in the medical device recall/correction notice of the Alere INRatio® Monitor System would be identified. The impact on the primary efficacy and principal safety endpoints would be evaluated in the population excluding the subset with these conditions."

Medical Conditions

The Urgent Medical Device Correction/Dear Healthcare Professional letter dated December 5, 2014 issued by Alere, Inc. listed a broad range of conditions in which the INRatio® PT/INR Monitor system should no longer be used (detailed in the background section above).

Definitions

Relevant medical condition – For the purposes of the sensitivity analyses, this term denotes diagnoses listed above. The conditions were subclassified as follows:

- Chronic – relevant medical conditions that are continuous or constant in nature

- Acute – relevant medical conditions that are time-limited

Recall-related time of observation – this is a censoring method used to precisely define the interval during which the Alere device performance issue may have exerted an effect on the INR values on a given patient in the ROCKET AF trial. The term "recall-related time of observation" considers not only the presence of relevant medical conditions and adverse event (AE)/serious adverse event (SAE) start and stop dates, but also the timing of INRs performed with the device in association with these intervals. In brief, the mere presence of a relevant medical condition is not sufficient to warrant exclusion from observation since it is the performance of an INR in association with this condition that is of probative value. For example, if a patient develops a relevant medical condition but an INR is not performed during that interval, the device recall would not be relevant. Conversely, INRs performed during a relevant medical condition may exert an effect even after the AE/SAE stop date, until a subsequent (i.e., unaffected) INR is performed.

Categorization of ROCKET AF subjects for the Sensitivity Analyses

It was noted that the conditions above listed could be found in four locations:

1) other relevant medical history,

2) adverse events,

3) serious adverse events, and

4) laboratory database.

Next, line listings of all coded (i.e., dictionary-derived) terms by body system or organ class were run for these 3 sets of conditions. These listings represented all possible conditions as reported by the investigators in these 3 eCRF locations.

Two physicians blinded to treatment assignments independently reviewed the 3 sets of outputs, identifying any conditions that matched those listed in the Alere letter. The two reviews were then compared and any discrepancies discussed and resolved. When agreement was reached and each physician was satisfied that all relevant conditions were identified, the terms in each list were divided into "acute" or "chronic".

For the terms found in "other relevant medical history" on the eCRF, all chronic conditions were used for censoring. For the categories of adverse event and serious adverse event, a different technique was employed. That is, for all terms deemed "acute", the investigator-reported start and stop date of the event would be used for censoring. For all terms deemed "chronic", the start date would indicate when censoring would begin but censoring would not end until the last observation.

Analysis Descriptions

The categorization described above applied to the population for efficacy and safety analyses regardless of whether patients received rivaroxaban or warfarin. For all of the analyses, the comparison is between rivaroxaban and warfarin, based on the ROCKET AF study. Descriptions of the 3 sensitivity analyses follow (see also Tables 3 & 4, below).

Sensitivity Analysis #1 – Efficacy

For this analysis, the outcome is the Primary Efficacy Endpoint (composite of all stroke and non-CNS systemic embolism). The trial population used is the intent-to-treat* (ITT) population (please refer to the foot note in Table 3) and the data scope is from randomization to the time of site notification (i.e., end of the trial). Patients who had a relevant chronic medical condition reported at baseline (i.e., in Relevant Medical History on the eCRF at the screening visit) are censored after the performance of their first INR of the study. For the remaining patients, the analysis also excludes specific time intervals if they had a relevant medical condition, with censoring as defined in "Recall-related time of observation" above.

Sensitivity Analysis #1 – Safety

For this analysis, the outcome is the Principal Safety Endpoint (composite of major and nonmajor clinically relevant bleeding). The trial population used is the Safety population and the data scope is on-treatment. Patients who had a relevant chronic medical condition reported at baseline (i.e., in Relevant Medical History on the eCRF at the screening visit) are censored after the performance of their first INR of the study. For the remaining patients, the analysis also excludes specific time intervals if they had a relevant medical condition, with censoring as defined in "Recall-related time of observation" above.

Sensitivity Analysis #2 – Efficacy

For this analysis, the outcome is the Primary Efficacy Endpoint. The trial population used is the ITT* population (refer to the foot note in table 3) and the data scope is from randomization to the time of site notification. This analysis completely excludes patients who had recall-related time of observation before the endpoint event or censoring. In other words, the only patients included in this analysis are those who did not have any recall-related time of observation before the endpoint event or censoring.

Sensitivity Analysis #2 – Safety

For this analysis, the outcome is the Principal Safety Endpoint. The trial population used is the Safety population, and the data scope is on-treatment. This analysis completely excludes patients who had recall-related time of observation before the endpoint event or censoring. In other words, the only patients included in this analysis are those who did not have any recall related time of observation before the endpoint event or censoring.

Sensitivity Analysis #3 – Efficacy

For this analysis, the outcome is the Primary Efficacy Endpoint. The trial population used is the ITT* population and the data scope is from randomization to the time of site notification. One can think of this analysis in terms of the occurrence of a primary efficacy event or a censoring event (i.e., reaching the end of the trial). For this analysis, if the primary efficacy event occurred or the end of the trial was reached outside of a recall-related time of observation, the patient is included. What occurred before the event is not considered, provided the event occurred outside of a period of recall-related time of observation. As such, this cohort of the ITT population is larger than that contemplated by Sensitivity Analysis #2 since it includes some patients who had recall-related time of observation.

Sensitivity Analysis #3 – Safety

For this analysis, the outcome is the Principal Safety Endpoint. The trial population used is the Safety population, and the data scope is on-treatment. As above, one can think of this analysis in terms of the occurrence of a principal safety endpoint or a censoring event (i.e., reaching the end of the treatment with study drug). For this analysis, if the principal safety endpoint occurred or the end of the treatment with study drug was reached outside of a recall related time of observation, the patient is included. What occurred before the event is not relevant, provided the event occurred outside of a period of recall-related

time of observation. As above, this cohort of the selected population is larger than that contemplated by Sensitivity Analysis #2 since it includes some patients who had recall-related time of observation.

In summary, 3 sensitivity analyses were designed to obtain a more thorough understanding of the potential effect of the recall on the results of the ROCKET AF study, each having its own strengths and limitations (Tables 3 & 4). Although they are labelled as Sensitivity Analysis 1, 2 and 3, the MAH considered them as having equal importance and provide complementary perspectives. Sensitivity Analysis #1 includes all patients but excludes specific recall-related times of observation; its focus is on exposure periods. The other 2 analyses focus on patients included or excluded from the analyses. Sensitivity Analysis #2 is the most conservative as it includes only those patients without any recall-related time of observation ahead of any endpoint event or censoring. Sensitivity Analysis #3 is less restrictive: a patient is included if an event or censoring occurred outside of a recall-related time of observation.

Table 3. Primary Efficacy: Overview of Sensitivity Analyses, Number of Patients Included for the Analyses of the Primary Efficacy Endpoint, and Total Time of Observation

Analysis	Abbreviated Description	Strengths	Limitations	Number of Subjects (%)		Total Time of Observation in Patient-years (%)	
				Rivaroxaban	Warfarin	Rivaroxaban	Warfarin
Original ROCKET AF				7081 (100)	7090 (100)	12689 (100)	12645 (100)
Sensitivity Analysis #1	**All patients but excludes specific recall-related times of observation**	Includes all available time of observation that is considered not impacted by the recall	Does not use all data after randomization as in the original ITT analysis Depends more than the other two sensitivity analyses on the definition of recall-related time of observation	7081 (100)	7090 (100)	11019 (86.8)	10978 (86.8)
Sensitivity Analysis #2	**Only those patients without any recall-related time of observation ahead of any endpoint event or censoring**	Most conservative analysis	Excludes the most subjects and data	4674 (66.0)	4672 (65.9)	8079 (63.7)	7993 (63.2)
Sensitivity Analysis #3	**A patient is included if an event or censoring occurred outside of a recall-related time of observation**	Uses the same time to event or censoring as in the original ITT analysis	The population is a subgroup of the original ITT population	6226 (87.9)	6221 (87.7)	11070 (87.2)	11029 (87.2)

Trial populations/Datascope: Efficacy = intent-to-treat* (ITT) [*excluding site 042012]/ randomization to the time of site notification (i.e., end of the trial)

Table 4. Principal Safety: Overview of Sensitivity Analyses, Number of Patients Included for the Analyses of the Principal Safety Endpoint, and Total Time of Observation

Analysis	Abbreviated Description	Strengths	Limitations	Number of Subjects (%)		Total Time of Observation in Patient-years (%)	
				Rivaroxaban	Warfarin	Rivaroxaban	Warfarin
Original ROCKET AF				7111 (100)	7125 (100)	9893 (100)	9979 (100)
Sensitivity Analysis #1	All patients but excludes specific recall-related times of observation	Includes all available time of observation that is considered not impacted by the recall	Depends more than the other two sensitivity analyses on the definition of recall-related time of observation	7111 (100)	7125 (100)	8466 (85.6)	8554 (85.7)
Sensitivity Analysis #2	Only those patients without any recall-related time of observation ahead of any endpoint event or censoring	Most conservative analysis	Excludes the most subjects and data	4926 (69.3)	4909 (68.9)	6411 (64.8)	6489 (65.0)
Sensitivity Analysis #3	A patient is included if an event or censoring occurred outside of a recall-related time of observation	Uses the same time to event or censoring as in the original analysis	The population is a subgroup of the original Safety population	5889 (82.8)	5893 (82.7)	8130 (82.2)	8246 (82.6)

Trial populations/Data scope: Safety = Safety/On treatment

Results of Sensitivity Analyses

Table 5 provides a summary of the subjects with recall-related time of observation while on Treatment. Approximately 1/3 in each treatment group had any recall-related time of observation. Total adverse events occurred in approximately ¼, with around 5% SAEs. Only 2.6-2.8% of subjects had a hematocrit measured to be <30% or >55%. Those with a chronic medical condition reported in the correction notice/recall made up just over 10%.

Table 5: Summary of Subjects With Recall-Related Time of Observation While on Treatment (up to Last Dose Plus 2 Days)
(Study 39039039AFL3001: Safety Analysis Set)

	Rivaroxaban (N=7111) n (%)	Warfarin (N=7125) n (%)	Total (N=14236) n (%)
Total Number of Subjects With			
Any Recall-Related Time of Observation	2449 (34.4)	2458 (34.5)	4907 (34.5)
Adverse Events	1785 (25.1)	1822 (25.6)	3607 (25.3)
Serious Adverse Events	368 (5.2)	385 (5.4)	753 (5.3)
Hematocrit (%) <0.3 or >0.55	201 (2.8)	187 (2.6)	388 (2.7)
Chronic Medical History	741 (10.4)	751 (10.5)	1492 (10.5)

Note: Percentages calculated with the number of subjects in each group as denominator.

Table 6 provides the output of the original ROCKET AF Efficacy analysis and the 3 sensitivity analyses performed. In the ITT population, rivaroxaban remained non-inferior to warfarin for the primary efficacy

outcome of composite of stroke and non-CNS systemic embolism. In the 3 sensitivity analyses performed, no meaningful shifts in the respective hazard ratios for rivaroxaban versus warfarin are observed and the confidence intervals did not appreciably widen. The upper limit of the 2-sided 95% confidence interval for each of the 3 sensitivity analyses remains well below the non-inferiority margin pre-specified in the original trial analysis of 1.46 in terms of risk (hazard) ratio, so that non-inferiority of the study drug is still secured and declared. Of the 3 sensitivity analyses, the largest difference in both event rate and hazard ratio was observed in Sensitivity Analysis 2, which included only those patients without any recall-related time of observation ahead of any endpoint event or censoring. Here the event rates per 100 patient-years increased slightly in both treatments compared to the original analysis (2.29 as opposed to 2.12/100 patient-years rivaroxaban, and 2.74 vs 2.42/100 patient-years warfarin), but the hazard ratio decreased modestly compared with the original hazard ratio, and the upper bound of the 95% CI remains well below the non-inferiority margin.

Table 6 Primary Efficacy: Treatment Comparisons of Rivaroxaban Versus Warfarin for the Primary Efficacy Endpoint (Adjudicated by CEC) (up to the Notification to the Site That the Required Primary Efficacy Endpoint Events Have Been Reached)

(Study 39039039AFL3001: Intent-to-Treat (Excluding SITE=042012) Analysis Set)

| Analyses | -------- Rivaroxaban -------- | | ---------- Warfarin --------- | | Rivaroxaban vs. Warfarin | |
| | N= 7081 | Event Rate | N= 7090 | Event Rate | Hazard Ratio | p-value |
	n/J	(100 Pt-yr)	n/J	(100 Pt-yr)	(95% CI) (a)	(b)
Original Trial Result	269/7081	2.12	306/7090	2.42	0.88 (0.74, 1.03)	0.117
Sensitivity Analysis 1	238/11019	2.16	268/10978	2.44	0.88 (0.74, 1.05)	0.169
Sensitivity Analysis 2	185/4674	2.29	219/4672	2.74	0.84 (0.69, 1.02)	0.079
Sensitivity Analysis 3	238/6226	2.15	268/6221	2.43	0.89 (0.74, 1.05)	0.174

Note: Primary Efficacy Endpoint is the composite of Stroke and Non-CNS Systemic Embolism.
Note: Event rate 100 pt-yr: number of events per 100 patient years of follow up.
Note: n = number of subjects with events, and J = number of subjects in each subgroup except for Sensitivity Analysis 1 where J= time of observation not related to the recall (in patient years).
Note: (a) For all analyses except Sensitivity Analysis 1, Hazard Ratio (95% CI) from the Cox proportional hazard model with treatment as a covariate. For Sensitivity Analysis 1, Hazard Ratio (95% CI) from the Cox proportional hazard model with treatment as a covariate, recall-related time of observation as a time-dependent covariate, and their interaction.
Note: (b) p-value (two-sided) for superiority of rivaroxaban versus warfarin in hazard ratio.

Table 7 provides the event rates and hazard ratios for the time to the first occurrence of bleeding events while on treatment in the Safety population for the original ROCKET AF Safety analysis, as well as for the 3 sensitivity analyses performed. Evaluation in the Safety population of the event rates and hazard ratios for the 3 sensitivity analyses showed that the bleeding event rates with rivaroxaban and warfarin for the principal safety endpoint remained comparable and little changed from the original analysis. In Sensitivity Analysis 2, which included only those patients without any recall-related time of observation ahead of any endpoint event or censoring, event rates increased somewhat on both rivaroxaban and warfarin compared to the original analysis (16.55 as opposed to 14.91/100 patient-years rivaroxaban, 15.38 as opposed to 14.52/100 patient-years warfarin) but the hazard ratio was only slightly higher, and the lower bound of the 95% CI in this comparison did not reach or cross the line of unity.

Table 7 Principal Safety: Treatment Comparisons of Rivaroxaban Versus Warfarin for the Principal Safety Endpoint (Adjudicated by CEC) While on Treatment (up to Last Dose Plus 2 Days)
(Study 39039039AFL3001: Safety Analysis Set)

Analyses	-------- Rivaroxaban --------		---------- Warfarin ---------		Rivaroxaban vs. Warfarin	
	N= 7111	Event Rate	N= 7125	Event Rate	Hazard Ratio	p-value
	n/J	(100 Pt-yr)	n/J	(100 Pt-yr)	(95% CI) (a)	(b)
Original Trial Result	1475/7111	14.91	1449/7125	14.52	1.03 (0.96, 1.11)	0.442
Sensitivity Analysis 1	1222/8466	14.43	1175/8554	13.74	1.05 (0.97, 1.14)	0.214
Sensitivity Analysis 2	1061/4926	16.55	998/4909	15.38	1.08 (0.99, 1.17)	0.092
Sensitivity Analysis 3	1222/5889	15.03	1175/5893	14.25	1.05 (0.97, 1.14)	0.191

Note: Principal Safety Endpoint is the composite of Major and Non-Major clinically relevant bleeding event.
Note: Event rate 100 pt-yr: number of events per 100 patient years of follow up.
Note: n = number of subjects with events, and J = number of subjects in each subgroup except for Sensitivity Analysis 1 where J= time of observation not related to the recall (in patient years).
Note: (a) For all analyses except Sensitivity Analysis 1, Hazard Ratio (95% CI) from the Cox proportional hazard model with treatment as a covariate. For Sensitivity Analysis 1, Hazard Ratio (95% CI) from the Cox proportional hazard model with treatment as a covariate, recall-related time of observation as a time-dependent covariate, and their interaction.
Note: (b) p-value (two-sided) for superiority of rivaroxaban versus warfarin in hazard ratio.

Conclusion

In summary, the MAH concluded that three sensitivity analyses support the original efficacy and safety analyses of the ROCKET AF trial and confirm the positive benefit/risk profile which rivaroxaban provides. The upper limit of the 2-sided 95% confidence interval for each of the three sensitivity analyses remains well below the non-inferiority margin pre-specified in the original trial analysis, allowing non-inferiority of the study drug to comparator to be still declared. In regard to the principal safety outcome of the composite of major and non-major clinically relevant bleeding, evaluation of the event rates and hazard ratios for the three sensitivity analyses, which in various ways took into consideration that the INRatio device was used to assess PT INR values in the trial, shows little change from the original analysis for rivaroxaban versus warfarin. The effect of potentially discrepant INR readings does not alter the conclusions of the ROCKET AF trial.

CHMP discussion on the MAH response

The Applicant has focused on the Rocket AF trial which is acceptable from an overall benefit risk balance perspective as this study was the pivotal study supporting the AF indication. The Japanese Rocket study had minimal impact on the CHMP conclusions on the AF application.

The CHMP considered that the identified deficiencies of the INR devices could potentially lead to inappropriate increases of the warfarin doses. However, this would probably have minor impact on the efficacy conclusions based on comparisons between the study arms as long as efficacy events during treatment are considered.

From a safety perspective, the major concern would be a potential for higher bleeding rates in the warfarin arms than would be expected in well managed warfarin therapy. This would mean that the initial conclusions reached by the CHMP when approving the NVAF indication, based on comparisons between the treatment arms, would potentially underestimate the bleeding tendency induced by rivaroxaban as compared to warfarin.

The Applicant has performed three sensitivity analyses based on the information provided by the INR device manufacturer (Alere) and taking both chronic and acute conditions into account. Patients who are judged to have had any recall related clinical condition or event are excluded in the second sensitivity analyses which resulted in an exclusion of approximately 31% of the patients. The first and third sensitivity analyses are focussing on the time periods when such events occurred and fewer patients are excluded in these analyses. In all three analyses only minor changes in the Hazard ratios between the rivaroxaban and warfarin study arms were observed as compared to the original analyses.

Overall, the Hazard Ratios of the efficacy results are unaffected by the sensitivity analyses.

For the principal safety endpoint, the event rates were higher in both groups in the sensitivity data sets and the Hazard ratio numerically tended to favor warfarin. For the sensitivity Analysis 2 (the most conservative analysis), warfarin was almost nominally statistically significantly better than rivaroxaban (p = 0.092, HR: 1.08 (CI: 0.99, 1.17). Among the analyses provided in this response this conservative analysis is considered by the CHMP to be the most relevant.

The analysis of the subcomponents of the principal safety endpoint supports the view that inaccurate INR measurements are not relevant for the overall safety conclusions. As can be seen in Table 2, the pattern of bleeding components was different between rivaroxaban and warfarin with death due to bleeding, critical organ bleeding and intracranial haemorrhage being nominally in favor of rivaroxaban. This differential pattern remained unchanged in the sensitivity analyses. The safety analyses results of ROCKET AF for the sensitivity analysis 2 is summarized in table 8 below. The results were consistent for sensitivity analysis 1 and 3 and not mentioned in this report.

Table 8: Safety analyses results of ROCKET AF for the sensitivity analysis 2

Treatment dosage	Xarelto 20 mg once a day (15 mg once a day in patients with moderate renal impairment) Event rate (100 pt-yr)	Warfarin titrated to a target INR of 2.5 (therapeutic range 2.0 to 3.0) Event rate (100 pt-yr)	Hazard ratio (95% CI) p-value
Major and non-major clinically relevant bleeding events	1061/4926 (16.55)	998/4909 (15.38)/	1.08 (0.99, 1.17) P=0.092/
Major bleeding events	266/4718 (3.90)	251/4702 (3.60)	1.08 (0.91, 1.28) p=0.382
Death due to bleeding*	18/4667 (0.26)	43/4670 (0.61)	0.43 (0.25, 0.74) p=0.002
Critical organ bleeding*	62/4670 (0.90)	100/4677 (1.43)	0.63 (0.46, 0.87) p=0.005
Intracranial haemorrhage*	37/4667 (0.54)	65/4671 (0.93)	0.58 (0.39, 0.87) p=0.008
Haemoglobin drop*	203/4711 (2.97)	153/4695 (2.19)	1.35 (1.10, 1.67) p=0.005
Transfusion of 2 or more units of packed red blood cells or whole blood	109/4694 (1.59)	87/4682 (1.24)	1.27 (0.96, 1.69) p=0.092
Non-major clinically relevant bleeding events**	/	/	/
All cause mortality	375/4669 (4.57)	392/4667 (4.81)	0.95 (0.82, 1.09) p=0.447

Nominally Significant
***Information on this parameter not available in Applicant's responses*

There was no indication that the most serious bleeding events on warfarin were related to recall related clinical conditions or events.

The CHMP considered that there are uncertainties related to these analyses, and it is assumed that the available information is probably somewhat limited e.g. fibrinogen levels have not been measured regularly and is rarely measured in clinical routine. The analyses were based on the assumption that INR

values are only affected in case of recall related conditions. This issue was further addressed during the procedure (see below). However, based on this assumption, the strategy chosen by the MAH is considered reasonable in order to identify patients that potentially had clinical conditions that would have affected the accuracy of the INR estimations according to the recall notice..

According to the MAH, it has been concluded by the manufacturer of the INR device (Alere INc) that less than 1% of the INR analyses would be affected by the identified deficiencies and that they would provide lower INR estimates than a correctly estimated ratio. During the regular TCs held with the MAH during this assessment, it became clear that plasma samples from the warfarin treated patients in the Rocket AF trial were drawn at week 12 and 24 for central laboratory prothrombin time (PT) testing. The primary intention of these analyses was to evaluate the potential impact of rivaroxaban on prothrombin time. The CHMP considered necessary to receive the data at week 6 and 12 as it should be possible to convert the PT values to INR values knowing the ISI (International Sensitivity Index) value for the reagents used even if this has to be done retrospectively. Useful comparisons could be made between the INR device values and INR lab values .provided that INR values were estimated with the device simultaneously (e.g. at the same day) as the plasma samples were drawn. Looking at the totality of such data, general trends for potential differences between INR values estimated by the two different methods could be identified among the warfarin treated patients. If such general trends exist, it could be further investigated if e.g. the sensitivity analyses 2 above were able to address differences that could be related to inadequate performance of the device.

The quality of the warfarin treatment was rather extensively discussed during the assessment of the application for prevention in AF. During this review, the concern was primarily that the TTR (Time in therapeutic range) was comparatively low with a median TTR of 58%. The concern was primarily related to INR values below the target range. This could be expected to result in a potential for reduced bleeding tendency and a risk for loss of efficacy during warfarin therapy. Thus, this tendency would on group level rather counteract a generally increased bleeding risk in the warfarin arm induced by inappropriately high warfarin doses.

The bleeding rates and the bleeding characteristics in the warfarin arm in the Rocket trial is considered to be within the expected range as described in published literature. The incidence of major and clinically relevant bleedings among the warfarin treated patients in the Rocket AF trial also compares well with the bleeding incidences observed in recently performed trials in AF populations with other oral anticoagulants (dabigatran, apixaban and edoxaban pivotal trials in atrial fibrillation). Thus, in summary, the available data do not indicate a generally higher bleeding rate in the warfarin arm of the Rocket AF trial than observed in similar recent trials or in the literature.

It is also agreed with the MAH that the incidences and the characteristics of bleedings induced by rivaroxaban in comparison with warfarin were characterised in the large clinical trials in other populations for secondary prevention of DVT/PE. The characteristics of the study populations in these studies are however similar to the Rocket AF study population and regarded as supportive for safety considerations since the doses of rivaroxaban were similar to those given in the Rocket AF trial. It can also be noted that in the pooled analyses of the DVT and PE studies, the major and clinically relevant bleedings were numerically fewer and major bleedings significantly fewer in the rivaroxaban arm (HR 0.93, 95% CI 0.81;1.06 and HR 0.55, CI 0.34, 0,79, respectively). The bleeding characteristics were similar to those observed in the Rocket AF trial. Thus, the CHMP concluded that there is external support for the relevance of the data derived from the Rocket AF trial.

In summary, the initial conclusions derived from the Rocket AF trial are supported by performed sensitivity analyses and by external evidence from large trials in other therapeutic areas comparing the

efficacy and safety of rivaroxaban vs warfarin. However, it was considered necessary to request additional analyses of the INR values provided by the two methods as follows:

1. Are there general trends for differences between INR values estimated by the two different methods among the warfarin treated patients? Please provide Bland-Altman plots of differences vs. means of the two methods with lines for upper and lower 95% confidence interval limits and the mean for the differences at week 12 and week 24.

2. To address that differences could be related to inadequate performance of the device, further exploratory analyses should be performed, e.g. by comparing the INR values provided with the two methods among the patients selected and excluded in the sensitivity analysis number 2.

3. To provide Bland-Altman plots (as under 1.), for (a) patients with and without any Recall-related Time of Observation separately and for (b) patients with and without impact by Recall-related Time of Observation at the measurement time points separately.

4. The preliminary data provided on INR values estimated simultaneously on weeks 12 and 24 with the two methods (POC device vs. laboratory testing) indicate that discrepancies of potential clinical relevance were rather frequently observed (approximately in 35% of the estimations). The majority of these discrepancies were related to a lower device INR value compared with Lab INRs. Cross tabulations are requested for the number of test results with Lab INR vs. POC INR for the following four categories: INR<2, 2-3, >3-4 and > 4.

5. For those patients who had discrepancies that can be expected to affect decision making, i.e, where the device INR was lower than the lab INR by one category, outcome data in terms of primary efficacy outcome and bleedings are requested. Similar calculations are requested for outcome data for patients where the POC INR was lower than the lab INR by two categories. These outcome data should be compared to the outcome data in patients that had measurements that were in concordance (i.e. that were in the same INR categories of below, within, and above the therapeutic 2-3 range).

6. A discussion and any additional analyses of potential value are requested to provide reassurance that the results and conclusions from the Rocket AF trial are still valid.

The MAH provided an integrated response to the above CHMP questions 1-3.

Summary of the MAH response Questions 1-3

Bland–Altman plots (Difference plots)

The Bland-Altman plot (Bland & Altman, 1986 and 1999) is a graphical method of data plotting to compare two measurements techniques. In this graphical method, the differences (or alternatively the ratios) between the two techniques are plotted against the averages of the two techniques or one of the two techniques.

The graph displays a scatter diagram of the differences (or ratios) plotted against the averages (or one) of the two measurements. Horizontal lines are drawn at the mean difference, and at the limits of agreement, which are defined as the mean difference plus and minus 1.96 times the standard deviation of the differences. Because the differences (ratios) between the Device based INR and Lab based INR for the ROCKET study do not appear to be symmetric, the graph will additionally display the horizontal line for the median difference.

The Bland-Altman plot method only defines the intervals of agreement, however it does not indicate whether those limits are acceptable or not. The acceptable limits can be defined based on the biological goals and clinical setting.

Orientation to the graphs

For each graph the X and Y axes are labelled with what is plotted. Each point on the plot represents a (x, y) calculated from a Warfarin subject's matched pair of Device based and Lab based INRs. The solid horizontal line represents the mean difference. The dashed horizontal line is the median difference of all points. The upper and lower dotted horizontal lines are mean difference + and -1.96 x Standard Deviation. N is the number of matched samples plotted. The numerical values for the Median, and Mean +/- 1.96 x Standard Deviation, are noted in the left lower corner of each plot. For the Device INR, any value that is at least 6.1 was reported as 6.1 in the database. For these subjects, the difference between the Device based INR of 6.1 and the Lab based INR do not represent the actual difference, except for subjects whose device values were actually 6.1, as there is no means to separate out these subjects in the database. Therefore, for all the plots, subjects with Device INR of 6.1 are excluded. With the X axis ranging from 0 to 14 and the Y axis ranging from -12.5 to 5 for absolute differences or -2.0 to 3.5 for the ratio of differences, isolated data points are not shown in the plots, but included in the calculations.

Analysis Descriptions

To address Request 1 and 2, the MAH   included 3 types of Bland-Altman plots, done separately for measurements at Week 12 and at Week 24. The 3 types of Bland-Altman plots are mentioned below and show:

- the difference of the Device INR minus the Lab measured INR against the average of the Device INR and Lab INR (i.e.,[Device based INR + Lab based INR]/2),

- the difference (Device based INR - Lab based INR) vs the Lab based INR,

- the ratio of the difference over Lab based INR ([Device based INR - Lab based INR] / Lab based INR) vs the Lab based INR. For example, an INR value of 2.0 for the device and 2.5 for the Lab is expressed as -0.20 (calculated as (2.0-2.5)/2.5)), corresponding to a 20% lower Device based INR value relative to the Lab based INR.

For each of these 3 types of Bland-Altman plots 10 different figures were generated corresponding to the Warfarin subjects to be assessed:

To address Request 1, plots are shown for all Warfarin subjects in the Safety Analysis set with Device based INR and Lab based INR at Week 12 (n=5702 subjects) and at Week 24 (n=5437 subjects).

GINR44A: Difference (Device - Lab) vs Average of Device INR and Lab INR at Week 12 for Warfarin Subjects; Safety Analysis Set (Study 39039039AFL3001)
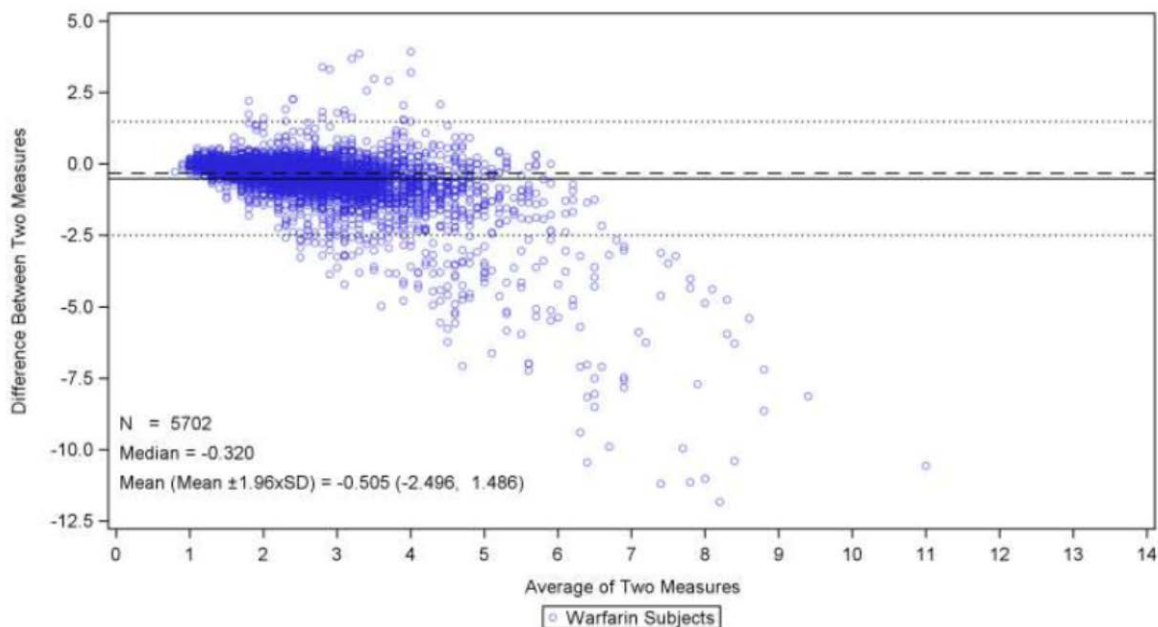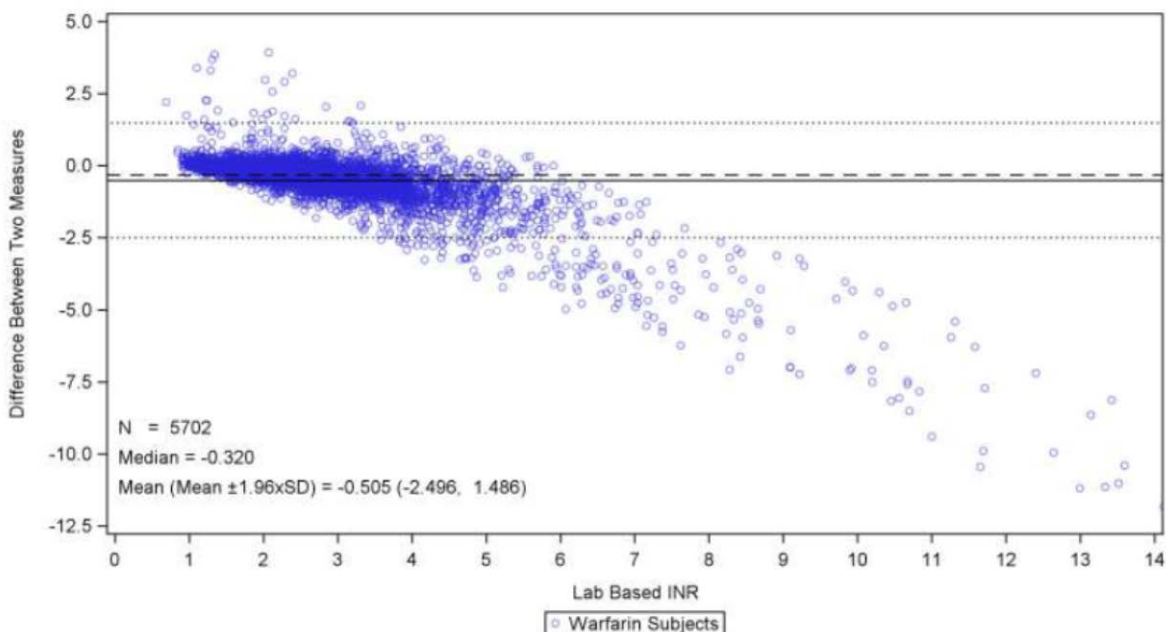

Notes:

Device based INR at least 6.1 is excluded.

The solid horizontal line is mean difference. The dashed horizontal line is median difference.

The upper and lower dotted horizontal lines are mean difference + and - 1.96 x Standard Deviation.

GINR74A: Difference (Device - Lab) vs Lab INR at Week 12 for Warfarin Subjects; Safety Analysis Set (Study 39039039AFL3001)

GINR84A: Ratio of Difference and Lab INR ([Device - Lab]/Lab) vs Lab INR at Week 12 for Warfarin Subjects;   Safety Analysis Set (Study 39039039AFL3001)
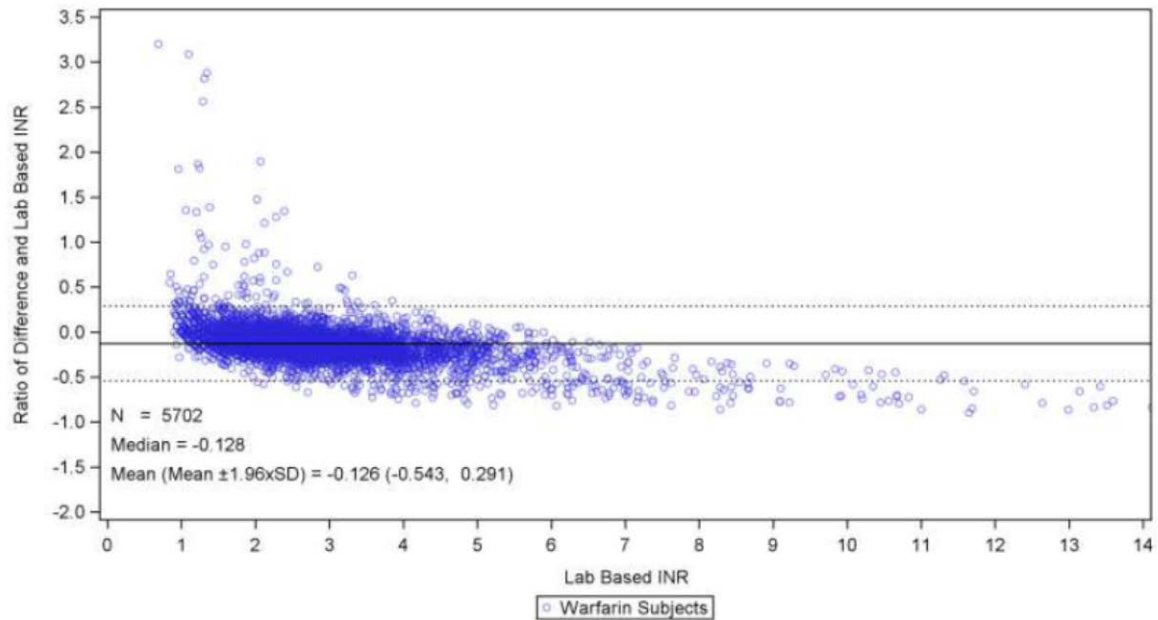
To address Request 2a, plots are shown in Warfarin subjects in the Safety Analysis set with recall related time of observation at any time of the study (n=2074 at Week 12, n=1988 at Week 24) and without recall related time of observation at any time of the study (n=3628 at Week 12, n=3449 at Week 24).

To address Request 2b, plots are shown in Warfarin subjects in the Safety Analysis set with recall related time of observation at the measurement time points (n=769 at Week 12, n=735 at Week 24) and without recall related time of observation at the measurement time points (n=4933 at Week 12, n=4702 at Week 24).

The Bland-Altman plots related to Requests 2a and 2b provide a picture which is very similar to the figures presented previously.

From these plots the MAH concludes the following:

- The pattern of absolute and relative differences between the Device based INR and Lab based INR looks similar for the population of subjects contributing measurements at Week 12 and the population of subjects contributing measurements at Week 24.

- On average the Device based INR measurements are lower than the Lab based INR values and the absolute differences tend to increase with increasing INR.

- For the absolute difference (Device INR minus the Lab INR) in all Warfarin subjects the median difference (-0.32 at Week 12, -0.31 at Week 24) is smaller than the (arithmetic) mean difference (-0.505 at Week 12, -0.513 at Week 24).

- Similar conclusions are reached when plotting the Lab based INR on the X axis rather than the average of the Device and Lab based INR measurement.

- With respect to the ratio of the difference over the Lab based INR median and (arithmetic) mean are of similar magnitude. Both show that on average the Device INR is about 13% lower than the Lab based INR. Visual inspection of the plots seems to suggest that this relative difference may not be constant across the entire range of Lab based INR values, but that the average relative

difference is slightly lower with lower Lab based INR and slightly higher with higher Lab based INR than the 13% average.

- The absolute and relative differences between Device and Lab based INR look largely similar for a) subjects with and without recall related time of observation at any time of the study and also b) for subjects with and without impact by recall-related time of observation at the INR measurement time point.

The MAH believes that the Bland-Altman plot series that compares the ratio of Difference (Device – Lab)/Lab INR vs Lab based INR (set #3), as opposed to the Difference (Device – Lab) vs Lab INR (set #2) or Difference (Device – Lab) vs Average of Device INR and Lab INR (set #1), is a clearer representation of the data, as it provides a relative, rather than absolute, comparison to the central Lab assay. Only small differences are seen in the mean or the median, both of which are similar in their values.

A simple graphical presentation of the relationship between device based INR and Lab INR was requested by the Rapporteurs and such a scatter plot was provided by the MAH as an attachment to the responses to the CHMP. Furthermore the attachment provided a discussion of the discrepancies between the two methods where the definition for a discrepancy was adapted from the 2007 ISO 17593 guidance document. This information is summarised below.

Lab Based INR vs Device Based INR at Week 12 for Subjects With or Without any Recall related Time of Observation; Safety Analysis Set
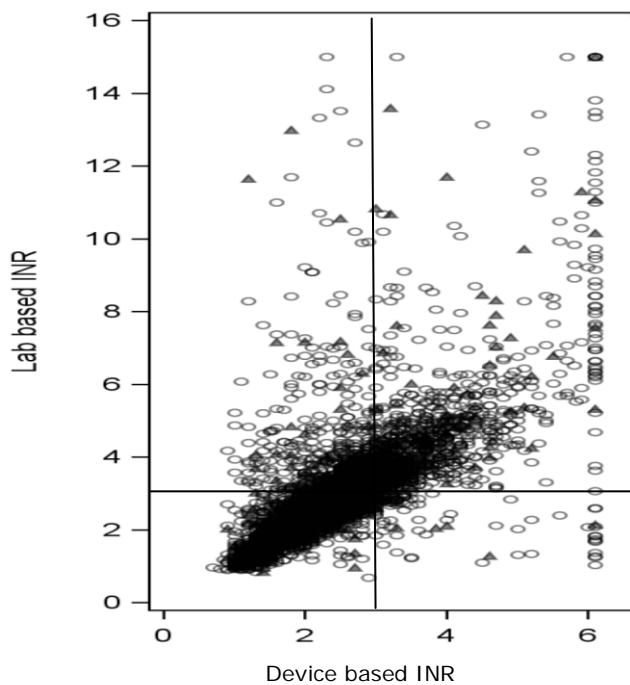
Device based INR
(circles-not impacted patients, triangles –impacted patients)

Table TSIINR26ZX [see below] displays the number and percent of subjects with discrepancy between device and lab based INR values With or Without any Recall-related Time of Observation at Week 12 or Week 24. Discrepancy is defined as a device INR that is outside of the corresponding lab based INR +/- 0.5 if the lab based INR < 2 OR a device based INR is outside of the corresponding lab based INR +/- 30% if lab based INR ≥ 2. These parameters were adapted from the 2007 ISO 17593 guidance document for INR interval 4.6 to 6.0. As no specific allowable difference is given by the 2007 ISO 17593 guidance document as this is seen as a supratherapeutic interval, the MAH expanded the +/- 30% rule for INR from 2 to 4.5 to INR beyond 4.5.

Table 3 — Performance criteria

INR interval	Allowable difference (90 % of results)	Allowable bias (average difference)
Below 2,0	± 0,5	NA
2,0 to 4,5	± 30 %	± 0,3
4,6 to 6,0	NA	NA
NA = not applicable.		

In table TSIINR26ZX, subjects with an unascertainable difference are not included in the Yes or No category; hence, the addition of the percents for N and Y do not total 100%. However, those subjects are included in the population size as 'n' when percentages in the table are calculated. No discrepancy was observed in almost 90% of those warfarin treated subjects. Subjects whose Device based INRs > 6 are not included.

TSIINR26ZX: Subjects With Discrepancy Between Lab Based INR and Device Based INR at Week 12 or at Week 24 for Subjects With or Without any Recall-related Time of Observation
(Study 39039039AFL3001:   Safety Analysis Set)

Subject Group	Week 12 (N=5766) n (%)	Week 24 (N=5507) n (%)
With Any Recall-related Time	775	745
Discrepancy	775	745
N	672 (86.7)	645 (86.6)
Y	97 (12.5)	90 (12.1)
Without Any Recall-related Time	4991	4762
Discrepancy	4991	4762
N	4280 (85.8)	4072 (85.5)
Y	653 (13.1)	630 (13.2)

In terms of differences for measurements With or Without recall related time, the data are consistent with previously provided information that nearly 90% of the Warfarin treated subjects in both groups had no discrepancy observed between the Lab based INR value and Device based INR value at Week 12 or at Week 24 (87% of INR fall in range).

CHMP discussion on questions 1-3

When analysing the MAH data, it appears that 64 INR values were excluded in these analyses as the device INR were ≥ 6.1. Taking the overall picture into account, it is probable that the majority of these values would also have had a high lab INR and they would thus have had less influence on the potentially clinically relevant differences leading to a decrease of potential clinical impact.

It is noted that the week 24 scatter plots are very similar to the week 12 figure and they are not provided in this report.

From a visual impression of the curves, it is agreed that the majority of the discrepancies were related to lower device based INRs as compared to lab INRs as further detailed below.

However, it needs to be highlighted that the potential impact of the discrepancies on clinical decision making cannot be easily estimated based on the Bland-Altman plots.

A pronounced similarity is observed between patients that had a recall related time of observation and the complete data set, irrespectively if the patients had a recall related time at any time of the study or at the measurement time points.

Therefore, this can question to some extent the relevance of the originally performed sensitivity analyses which were focussing on identification of patients that had a device recall time of observation (see above)

The targeted therapeutic range is INR between 2 and 3. If the Lab based INRs are regarded as "true values" the upper left rectangle in the scatter plot above representing the relation between Lab INR and device INR would represent situations where a dose reduction from a safety perspective would potentially be worth considering. Under such an assumption this signal was not caught by the device. The plot provides an additional visual picture of the observation that most of the discrepancies reported were related to a lower device INR as compared to the lab INR. Of note the 24 w scatter plot is similar to the 12 w plot. It should be recognised that many of the discrepancies observed were small although above (lab) and below (device) INR 3, respectively. An adjustment of the long term warfarin dose due to a value outside the therapeutic range should normally not be based on a single value but rather on repeated close measurements possibly after a short term dose amendment.

The statement that the ratio Difference (Device – Lab)/Lab INR vs Lab based INR (set #3) would be a clearer representation of the data, as opposed to set #1 and 2, appears questionable from a clinical relevance perspective.

CHMP overall conclusions Q1-3

In conclusion, the applicant has provided the requested data and an appropriate mathematical description of the concordance of the two INR estimation methods. Overall, the CHMP concluded that the majority of discrepancies observed with the INR device were related to a lower device INR value as compared to the lab INR value. It must however be emphasised that the provided INR laboratory data only reflects the INR results at two specific time points (week 12 and week 24) and that clinical decisions on dosing should not be based on single INR measurements.

It is also concluded that the originally performed sensitivity analyses provided in October 2015 do not seem to have captured the discrepancies observed at week 12 and 24. These analyses were based on retrospective identification of patients with clinical or laboratory conditions where the device estimated INR would be unreliable according to the Alere recall notice information.

With the criteria set up in the 2007 ISO 17593 standards, the majority of samples were found to be concordant. These criteria are however somewhat arbitrary and cannot easily be interpreted in terms of relevance for clinical decisions on dosing. Further analysis and assessment to elucidate the potential clinical relevance of these data are made in the following questions.

Question 4

The preliminary data provided on INR values estimated simultaneously on weeks 12 and 24 with the two methods (POC device vs. laboratory testing) indicate that discrepancies of potential clinical relevance were rather frequently observed (approximately in 35% of the estimations). The majority of these discrepancies were related to a lower device INR value compared with Lab INRs. Cross tabulations are requested for the number of test results with Lab INR vs. POC INR for the following four categories: INR<2, 2-3, >3-4 and > 4.

Summary of the MAH response Q4

Cross tabulations of Device based vs Lab based INR, defining the therapeutic range as INR values from 2 to 3, are provided below using the INR categories <2, 2-3, >3-4 and >4 as cut offs. Device based INR measurements reported as 6.1 (any Device INR of at least 6.1 was reported as 6.1 in the database) are included in the table resulting in 5766 subjects With Device based INR and Lab based INR measurements at Week 12 and 5507 subjects at Week 24.

Table 1: TSIINR23ZYa: Cross Tabulation of Lab INR vs Device Based INR at Week 12 for Warfarin Subjects
(Study 39039039AFL3001:   Safety Analysis Set)

Visit Name: WEEK 12

	--- Lab INR ---				
	<2	2 - 3	>3 - 4	>4	Total
	n (%)	n (%)	n (%)	n (%)	n (%)
Device INR	1356 (23.5)	2405 (41.7)	1238 (21.5)	767 (13.3)	5766 (100)
<2	1239 (21.5)	604 (10.5)	54 (0.9)	47 (0.8)	1944 (33.7)
2 – 3	96 (1.7)	1740 (30.2)	793 (13.8)	172 (3.0)	2801 (48.6)
>3 – 4	8 (0.1)	50 (0.9)	358 (6.2)	291 (5.0)	707 (12.3)
>4	13 (0.2)	11 (0.2)	33 (0.6)	257 (4.5)	314 (5.4)

Note: Percentages calculated with the total number of subjects.

The results of Table 1 TSIINR23ZYa are similar to those observed from the Bland-Altman plots, showing that the Device INR yields overall lower values than the Lab INR.

62% (3594 out of 5766) of the measurements were in the same INR categories of below (<2), within (2-3), and above (>3 - 4 and >4) the therapeutic 2-3 range; Group 1;

4% (211 out of 5766) of the measurements had higher Device INR values compared with Lab INRs (according to the INR categories); Group 2;

34% (1961 out of 5766) of the measurements had lower Device INR values compared with Lab INRs. For at least some of the measurements where both the Device and the Lab INR are above the therapeutic range (e.g. Device INR is >3 – 4 and Lab INR is >4; 291 measurements) it is likely that the clinical action would be similar (according to the INR categories); Group 3.

Only 5% (273 out of 5766) of the measurements for Device INR were lower by two categories compared to the Lab INRs, meaning the dose would have been increased or maintained when should have been decreased (according to the INR categories).

The concordance (Group 1) / absence of concordance (Group 2 and 3) in Device vs Lab INR measurements was used to define patient subpopulations in whom efficacy and safety outcome data were explored, hypothesizing that a lack of concordance could have affected clinical decision making (for details please refer to request 5).

Similar findings are present in at Week 24. It is of note that the subjects do not always stay in their category throughout the study as they were in Week 12 or Week 24.

CHMP discussion on Question 4

The requested data have been provided by the MAH and it is agreed that they are providing additional support to confirm the overall conclusions derived from the Bland-Altman plots. By excluding the 291 measurements where Device INR is >3 – 4 and Lab INR is >4 , the proportion of measurements with a lower Device INR values compared with Lab INRs values are reduced from 34% to 29% according to the rapporteur's calculation, a number which still could be of some concern in relation for the possibilities of inappropriate dosing.

Questions 5 and 6

For those patients who had discrepancies that can be expected to affect decision making, i.e, where the device INR was lower than the lab INR by one category, outcome data in terms of primary efficacy

outcome and bleedings are requested. Similar calculations are requested for outcome data for patients where the POC INR was lower than the lab INR by two categories. These outcome data should be compared to the outcome data in patients that had measurements that were in concordance (i.e. that were in the same INR categories of below, within, and above the therapeutic 2-3 range).

Summary of the MAH responses Q 5 and 6.

Therefore the following analyses were conducted and have been summarized below

- Analysis 1: Efficacy and safety of Warfarin subjects in whom overall there was a discrepancy in the distribution of the Device based INR values in categories <2, 2-3, >3-4 and >4 relative to Lab based INR, compared with subjects Without a Discrepancy. (A Discrepancy is defined as Device based INR being outside of Lab based INR +/- 0.5 if Lab based INR < 2, and Device based INR is outside of Lab based INR +/- 30% if Lab based INR $\geq$ 2.) This analysis refer back to comparison of Device based INR to Lab based INR provided on Nov 13, 2015 (Analysis 2: Efficacy and safety of Warfarin subjects where the measurements were in Concordance (in the same INR categories of below (<2), within (2-3), and above (>3 - 4 and >4) the therapeutic 2-3 range).

- Analysis 3: Efficacy and safety of Warfarin subjects where measurements had lower Device based INR values by (at least) one category compared with Lab INRs..

- Analysis 4: Efficacy and safety of Warfarin subjects where measurements for Device based INR were lower by 2 categories (>1 INR unit) from the Lab INRs..

- Analysis 5: Efficacy and safety of Warfarin subjects without Paired Device INR and Lab INR. Note: Analysis 5 was performed to put Analysis 1-4 into perspective as a significant number of subjects and events from the overall Warfarin group does not contribute to these Paired Device vs. Lab analysis sets.

For example, a subject with Device INR of 2.9 and Lab INR of 3.2 would be considered as "Without Discrepancy" in Analysis 1, as "Without Concordance" in Analysis 2, as "With Lower Device INR by One Category" in Analysis 3, as "Without Lower Device INR by Two Categories" in Analysis 4 and not considered at all in Analysis 5.

These analyses were done for affected subjects in Week 12 or Week 24 as well as for subjects affected in Week 12 and Week 24.

<u>Results of the Primary Efficacy and Principal Safety Event Rates of Warfarin Subjects</u>

Analysis 1: Efficacy and safety of Warfarin subjects in whom overall there was a Discrepancy in the distribution of the Device based INR values relative to Lab based INR, compared with subjects Without a Discrepancy.

- The primary efficacy event rates were higher for subjects With Discrepancy compared with those Without Discrepancy at Week 12 (2.53/100 pt-yrs vs 1.96/100 pts-yrs), Week 24 (2.5/100 pt-yrs vs 1.56/100 pts-yrs), and at Both Week 12 & 24 (3.62/100 pt-yrs vs 1.57/100 pts-yrs). As the device reportedly more commonly underreads the INR value, potentially leading to increases in warfarin dose, it is expected that the efficacy event rates would be lower in the With Discrepancy groups, but the opposite was observed.

- The principal safety endpoint event rates were higher for subjects With Discrepancy compared with those Without Discrepancy at Week 12 (15.11/100 pt-yrs vs 13.04/100 pts-yrs), Week 24 (14.96/100 pt-yrs vs 12.71/100 pts-yrs), and at Both Week 12 & 24 (19.85/100 pt-yrs vs 12.32/100 pts-yrs). This could be expected if up-titration of warfarin dosing were performed to address Device INRs below the therapeutic range in subjects With Discrepancy.

Analysis 2: Efficacy and safety of Warfarin subjects where the measurements were in Concordance (in the same INR categories of below (<2), within (2-3), and above (>3 - 4 and >4) the therapeutic 2-3 range).

- While the primary efficacy event rates for subjects With and Without Concordance at Week 12 are almost identical (2.05/100 pt-yrs vs 2.00/100 pts-yrs), the efficacy event rates for subjects With and Without Concordance at Week 24 (1.59/100 pt-yrs vs 1.84/100 pts-yrs) increase and With and Without Concordance at Both Week 12 & 24 (1.73/100 pt-yrs vs 1.59/100 pts-yrs) decrease, meaning the trend is not similar. It might be expected that the primary efficacy event rates for subjects With Concordance at Week 12, Week 24, and Both Week 12 & 24 should be nearly identical, this was not observed, suggesting that the populations at the 2 time points are somewhat different, substantiating the limitation that these single time point measurements during a long term study cannot provide a full reflection of the duration of treatments of patients in the entire study. As the device reportedly more commonly underreads the INR value, potentially leading to a higher warfarin dose when not necessarily needed, it might be expected that the event rates would be lower in the Without Concordance groups, but this was only seen in the Both Week 12& 24 group.

- More of a disparity is also observed for the principal safety endpoint. The principal safety endpoint event rates for subjects With and Without Concordance at Week 12 and at Week 24 are dissimilar (12.88/100 pt-yrs vs 14.01/100 pts-yrs, and 12.60/100 pt-yrs vs 13.63/100 pts-yrs, respectively), although closer With and Without Concordance at Both Week 12 & 24 (12.20/100 pt-yrs vs 12.86/100 pts-yrs). Again, as the device reportedly more commonly underreads the INR value, potentially leading to higher warfarin dose when not necessarily needed, here this is in keeping with that postulate.

Analysis 3: Efficacy and safety of Warfarin subjects where measurements had lower Device based INR values by (at least) one category compared with Lab INRs.

- Likewise to Analysis 2, the primary efficacy event rates in Analysis 3 for subjects With and Without Lower Device INR by One Category at Week 12 are almost identical (2.06/100 pt-yrs vs 2.02/100 pts-yrs), but the event rates for subjects With and Without Lower Device INR by One Category at Week 24 (1.89/100 pt-yrs vs 1.57/100 pts-yrs) and With and Without Lower Device INR by One Category at Both Week 12 & Week 24 (1.97/100 pt-yrs vs 1.59/100 pts-yrs) decrease, meaning they are not similar. It could be expected that the primary efficacy event rates for subjects Without Lower Device INR by One Category at Week 12, Week 24, and Both Week 12 & 24 should be nearly identical, this also was not observed, again suggesting that the populations at the 2 time points are somewhat different. If an increase in warfarin doses was prescribed to those in the With Lower Device INR by One Category group, it would be expected to see lower efficacy rates, but this was not seen at either time point, or at Both Week 12 & 24.

- In Analysis 3, the principal safety endpoint event rates are somewhat higher for subjects With as opposed to subjects Without Lower Device INR by One Category at Week 12 (14.48/100 pt-yrs vs 12.70/100 pts-yrs, Week 24 (13.93/100 pt-yrs vs 12.50/100 pts-yrs), and Both Week 12 & 24 (14.63/100 pt-yrs vs 12.21/100 pts-yrs). Again, this could be viewed as in keeping with the postulate that warfarin may have been dosed higher if the Device INR was lower than the Lab INR.

Analysis 4: Efficacy and safety of Warfarin subjects where measurements for Device based INR were lower by 2 categories (>1 INR unit) from the Lab INRs.

- The differences in primary efficacy event rates are somewhat greater in Analysis 4 than in Analysis 3 for subjects With and Without Lower Device INR by Two Categories at Week 12

(2.50/100 pt-yrs vs 2.01/100 pts-yrs), With and Without Lower Device INR by Two Categories at Week 24 (2.05/100 pt-yrs vs 1.66/100 pts-yrs), and With and Without Lower Device INR by Two Categories at Both Week 12 and Week 24 (1.33/100 pt-yrs vs 1.65/100 pts-yrs), but only in this very last comparison does the difference reach >0.25/100 pt-yrs, triggered by only 1 event in the With Lower Device INR by Two Categories group. Again, it might be expected that the primary efficacy event rates for subjects Without Lower Device INR by Two Categories at Week 12, Week 24, and Both Week 12 & 24 should be essentially the same, but this was only the case at Both Week 12 & 24. If an increase in warfarin doses was prescribed to those in the With Lower Device INR by Two Categories group, it might be expected to see lower efficacy rates, but actually higher primary efficacy event rates were seen at Week 12 or at Week 24.

- The principal safety endpoint event rates in Analysis 4 are higher for subjects With as opposed to subjects Without Lower Device INR by Two Categories at Week 12 (16.59/100 pt-yrs vs 13.15/100 pts-yrs, Week 24 (14.60/100 pt-yrs vs 12.91/100 pts-yrs), and Both Week 12 & 24 (21.33/100 pt-yrs vs 12.52/100 pts-yrs). These observed event rates suggest warfarin may have been dosed higher in the subjects With Lower Device INR by Two Categories.

Analysis 5: Efficacy and safety of Warfarin subjects Without Paired Device INR and Lab INR.

- Warfarin subjects Without Paired Device INR and Lab INR (1359 for Week 12, 1618 for Week 24 and 2077 for Week 12&24) represent a large portion of the total Warfarin Cohort (7125 subjects).

- The primary efficacy endpoint event rates subjects on warfarin Without Paired Device INR and Lab INR at Week 12 (4.37/100 pt-yrs), Week 24 (5.56/100 pt-yrs), and at Both Week 12 and Week 24 (4.63/100 pt-yrs) are higher than observed for subjects on warfarin for the ROCKET AF trial overall (2.42/100 pt-yrs) and also higher in the subjects group considered as Without Concordance or With Discrepancy.

- The principal safety endpoint event rates for subjects on warfarin Without Paired Device INR and Lab INR at Week 12 (23.44/100 pt-yrs), Week 24 (24.98/100 pt-yrs), and at Both Week 12 and Week 24 (22.36/100 pt-yrs) are higher than for the overall trial for subjects on warfarin (14.52/100 pt-yr) and also higher in the subjects group considered as Without Concordance or With Discrepancy.

- Comparing these events rates with the events rates seen in Analysis 2 - 4 indicates again that the population With and Without Paired Device INR and Lab INR are clearly different. It can be concluded that the analysis of the subset of patients for which a paired analysis 1-4 has been performed is not representative for the overall results in the total Warfarin group.

The Applicant has provided a number of points to Consider in Interpretation of the Analysis 1 – 5:

There are several limitations in interpreting the event rates, including the following confounding factors, selection biases and unquantifiable relations:

- Subjects are grouped based only on the patterns of Device INR vs Lab INR at 1 or 2 time points (i.e., Week 12 and/or Week 24). The subjects at Week 12 and/or Week 24 may not have the same type of Concordance patterns throughout the entire duration of the trial.

    o There is no clearly quantifiable relationship between the patterns at Week 12 and/or Week 24 and the event rates throughout the trial.

    o As a case in point, subjects do not have consistent patterns when comparing Week 12 and Week 24. For example, the number of subjects having the same pattern for Both Week 12 and Week 24 is substantially smaller than that for Week 12 alone and that for Week 24

alone. For example: In the Safety Analysis Set among 273 subjects with Lower Device INR by two Categories at Week 12 (309 for Week 24) only 39 subjects are in that category at Week 12 and Week 24.

- The pattern of Device INR vs Lab INR themselves do not have clearly quantifiable implications on dose adjustments, not at Week 12 nor at Week 24. The implication on dose adjustment throughout the study is elusive as well.

- The selection of subjects based on whether they have Paired Device and Lab INR at Week 12 and/or Week 24 introduces subject selection biases.

  - For example, subjects with primary efficacy endpoint events before Week 12 or Week 24 discontinued the study drug early, and thus did not have Paired Device INR and Lab INR. These early events enriched the subgroup of subjects Without Paired Device INR and Lab INR, leading to their artificially enlarged event rates. On the other hand, these early events reduced the numbers of events remaining for the subgroup of subjects With Paired Device INR and Lab INR, leading to their artificially reduced event rates.

- There are possible differences in known and unknown baseline variables or post-baseline variables (e.g., distributions of Lab INR at Week 12 and/or 24) among the subgroups of subjects. The corresponding event rates for these subgroups may be confounded with the differences in these variables.

- There can be severe multiplicity of spurious results.

Additional exploratory analyses

A number of additional exploratory analyses taking into consideration the cross tabulations proposed by CHMP were undertaken by the MAH. The goal was to better understand the INR results obtained with the Alere INRatio device in the ROCKET AF trial, their correlation with the laboratory based INRs performed at weeks 12 and 24 as part of a pharmacodynamics study, and the impact of these correlations on the key efficacy and safety outcomes of the ROCKET AF trial. These analyses were refined over time based on new information and insights as the analyses unfolded. From the MAH perspective the latest analysis (Version IV) is the most informative but in the spirit of full transparency, all four versions of the analyses were submitted with appropriate explanation on how they evolved.

In version IV rivaroxaban and warfarin treated patients that had paired INR samples were included. The Consistency/Inconsistency/Discrepancy criteria were defined as in the previous analyses and they were applied to both treatment groups. Comparisons with patients without paired samples were also made.

Review of the Version analysis set unexpectedly revealed that inconsistencies in INR measurements were associated with differences in rivaroxaban event rates as well as warfarin event rates, see figure below. Since the dose of rivaroxaban was not modified in the trial, the results could represent a chance finding, or alternatively, these data suggest that a discrepancy in INR measurements between lab and device is associated with differences in events rates in the subject population independent of dose modification. For the principal safety endpoint in particular, event rates in the Rivaroxaban group without discrepancy/with concordance tracked closely with event rates in Warfarin subjects without discrepancy/with concordance. At the same time, event rates in the Rivaroxaban group with discrepancy/without concordance were higher than the without discrepancy subjects and tracked closely with event rates in Warfarin subjects with discrepancy/without concordance.

Version IV suggests that observed differences in warfarin event rates cannot be attributed to the adjustments made in warfarin dose based on discrepancy-related device INRs.

Version IV analyses

GSIBL402W12: Kaplan-Meier Plots of Time From the First Occurrence of Principal Safety Endpoint (Adjudicated by CEC) While On Treatment (up to Last Dose Plus 2 Days) for Subjects With or Without Concordance Device INR and Lab INR at Week 12 (Study 39039039AFL3001: Safety Analysis Set)
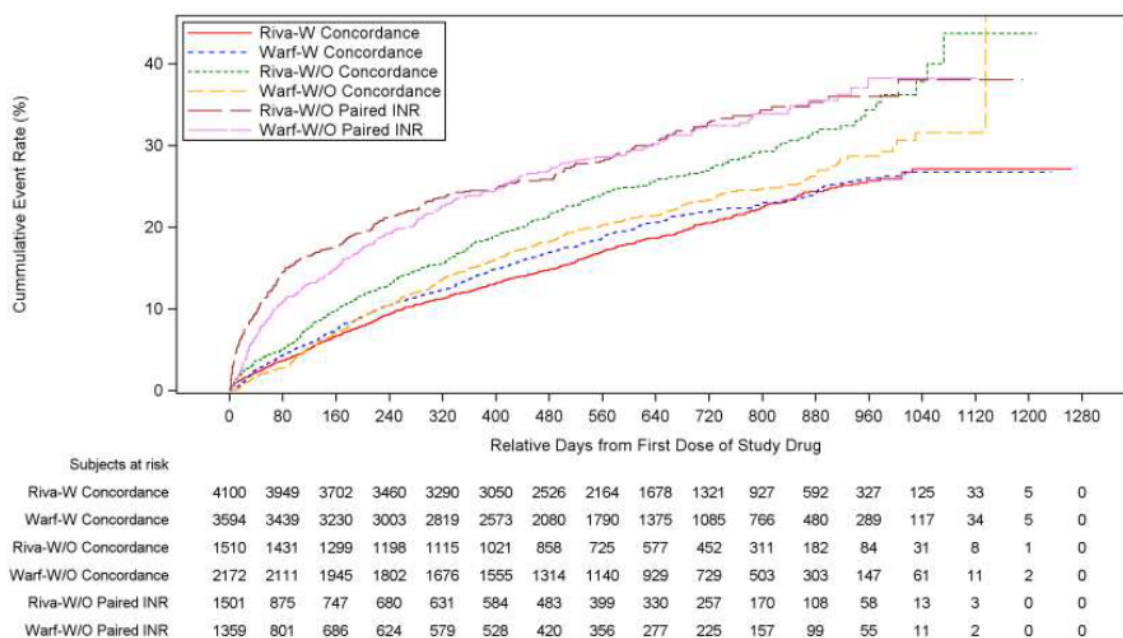
The following can be drawn from the Kaplan-Meier plots above. In the subpopulation without paired INR samples, the event rates are higher than in those with paired INR samples at Week 12 and/or Week 24, for both the Warfarin and Rivaroxaban subjects.

The absence of concordance between Device and Lab INRs is also associated with higher event rates in warfarin and rivaroxaban subjects which suggest that observed differences in warfarin event rates cannot be attributed to the adjustments made in warfarin dose based on discrepancy-related device INRs.

As discordant INR results appeared to be weakly associated with the main safety outcomes in the performed sensitivity analyses the following calculations were requested by the CHMP in order to further quantify the theoretical consequences of this level of increased risk:

- Calculate "population attributable risk" in the warfarin treatment arm of the study population based on 34% expected to have discordant INR results and the highest increased risk as estimated in "Analysis 3" of safety outcomes (Both week 12 & 24 increase from 12.21 to 14.63 / 100 PY).

- Apply the result to explore to what extent the event rates for the warfarin group in the table for safety analyses results as summarized in the SmPC (Table 2) would be reduced

- Calculate corresponding risk ratios comparable to the available hazard ratios (HR).

The requested analyses are summarised in the table 9 below where the two columns at right provide the modified RR and the corresponding Rate ratios.

Table 9

Study population	Patients with non-valvular atrial fibrillation					
	Original analyses				Analyses based on modified event rate	
Treatment dosage	Xarelto Event rate (100 pt-yr)	Warfarin Event rate (100 pt-yr)	HR (95% CI) p-value	Rate ratio	Warfarin modified event rate (100 pt-yr)*	Rate Ratio Xarelto vs Modified event rate
Major and non-major clinically relevant bleeding events	1,475 (14.91)	1,449 (14.52)	1.03 (0.96 - 1.11) 0.442	1.03	13.60	1.10
Major bleeding events	395 (3.60)	386 (3.45)	1.04 (0.90 - 1.20) 0.576	1.04	3.33	1.08
Death due to bleeding	27 (0.24)	55 (0.48)	0.50 (0.31 - 0.79) 0.003	0.50	0.44	0.55
Critical organ bleeding*	91 (0.82)	133 (1.18)	0.69 (0.53 - 0.91) 0.007	0.69	1.16	0.71
Intracranial haemorrhage*	55 (0.49)	84 (0.74)	0.67 (0.47 - 0.93) 0.019	0.66	0.74	0.66
Haemoglobin drop*	305 (2.77)	254 (2.26)	1.22 (1.03 - 1.44) 0.019	1.23	2.16	1.29
Transfusion of 2 or more units of packed red blood cells or whole blood*	183 (1.65)	149 (1.32)	1.25 (1.01 - 1.55) 0.044	1.25	1.12	1.48
Non-major clinically relevant bleeding events	1,185 (11.80)	1,151 (11.37)	1.04 (0.96 - 1.13) 0.345	1.04	10.53	1.12

*Modified Warfarin event rate based on "population attributable risk approach" assuming 34% subjects with discordant Device vs Lab based INR and relative bleeding risk increases derived for criterion of Device INR value lower than the Lab INR by at least one category (Analysis 3) in subjects with paired INR samples at both Week 12 and Week 24

CHMP discussion on questions 5 and 6

In this discussion part, each analysis is discussed individually prior to the overall CHMP conclusion.

In analysis 1 the definition of discrepancy is based on the 2007 ISO 17593 guidance document. The efficacy event rates were actually higher in the discrepancy group, which could reflect that these non-randomised groups had different characteristics and were at different risks. Another contributing possibility could be that treatment management after bleeding would be difficult in this study population at high embolic risk. In analysis 2, the concordance/lack of concordance is based on the four INR groups as defined in question 5. This can be regarded as an important analysis from a clinical perspective since a lack of concordance according to these criteria would lead to different signals and subsequent decision for a need of dose adjustment.

It is concluded that the differences between the concordant and non-concordant groups in outcome are rather small. This is considered reassuring in relation to potential clinical consequences of a INR device failure. The assumption that the patients in these different non-randomized groups have different

characteristics is further supported by these analyses, taking the overall picture with inclusion of efficacy data into account.

The analyses 3 and 4 can partly be regarded as subgroups to the groups compared in analysis 2. Analysis 4 reflects a more extreme situation with a pronounced lack of concordance and here the results can support an assumption that subjects with a lower device INR by two categories were given higher warfarin doses than needed. As discussed above the somewhat higher efficacy event rates among those with lower device INR may reflect that the populations are different but also that the handling of patients after a bleeding episode is difficult and may lead to efficacy failure in a high risk population. However, the groups are small in analysis 4 making any conclusions uncertain and the results in this subgroup appear to have affected the overall results to a limited extent as could be observed in analysis 2.

The higher efficacy and safety event rates in patients without paired device INR measurements may reflect different characteristics and risk profiles of the patients selected for paired measurements from those that were not selected. That patients without paired measurements could have had an efficacy event before w 12 may also contribute to the differences observed. The points to Consider in Interpretation of the Analysis 1 – 5: and related uncertainties are agreed to. However, while keeping in mind the uncertainties, it is of utmost importance to assess and better understand the potential impact of inaccurate estimations of the INR values with regards to the efficacy and safety results in the control arm of the Rocket AF trial.

The additional exploratory analyses requested by the CHMP included data from the rivaroxaban treated patients which was not the case in all the other previously discussed sensitivity analyses. Only version IV is discussed in this report as it is agreed that this version is the most relevant one. In conclusion, the additional exploratory analyses performed, including also the rivaroxaban patients, appears to support the conclusion that the population characteristics with and without concordance/discrepancy differ and that the differences observed are not only related to potentially inappropriate dosing of the warfarin patients but to other factors.

The efficacy data in the version IV analyses are not discussed by the MAH) but it can be noted from the data provided that the primary efficacy outcome was lower among the rivaroxaban treated patients with discrepancy at week 12 than in those without discrepancy (1.63 vs 2.53 per 100 pt-years, respectively). The same pattern is observed for those patients with and without discrepancies at week 24. This data could possibly reflect greater difficulties to adjust doses of warfarin than to adjust rivaroxaban doses after a bleeding episode.

The pattern observed could support a conclusion that inaccurate device estimations can have led to a somewhat increased bleeding tendency among the warfarin treated patients, however it could also theorically suggest an advantage for treatment with rivaroxaban in such situations.

The additional sensitivity analyses requested by the CHMP were intended to quantify the potential impact of discordant INR results. The review of these sensitivity analyses resulted in rather small differences in Hazard Ratios based on the modified event rates in the warfarin arm as displayed in Table 7 above.

The limitations of these analyses are fully acknowledged however they are considered to provide further reassurance for the relevance of the originally provided data assessed during the initial review of the Rocket AF trial.

2.3. MAH conclusions on the scientific analyses provided

The MAH would like to highlight that the ROCKET AF trial was not designed to validate the performance of the POC device or to calibrate it against a Lab based INR. Rather, the goal was to provide a blinded uniform standard throughout the trial to guide the investigators. To maintain the blind a POC INR measuring device was utilized for two reasons: 1) to capitalize on their capability to provide encryption of each INR value obtained and thereby improve the shamming and blinding capability for the trial, and 2) to facilitate more rapid and real time turnaround of INR values. The device chosen was CE marked and had 510k approval.

While the mean treatment duration of the ROCKET AF trial was 580 days in the Warfarin group with the majority of subjects receiving treatment for at least 18 months, only for 2 time points (Week 12 and Week 24) were trial Lab based INR values with a corresponding Device based INR value available. A device performance analysis was performed in this subset of patients at Week 12 and Week 24. It was revealed that there was a linear correlation for the Lab based INR value to the Device based INR value at Week 12 and 24 for the subjects With or Without any recall-related time of observation. Using discrepancy specifications adapted from those defined in 2007 ISO 17593, nearly 90% of the Warfarin treated subjects in the ROCKET AF trial showed no discrepancy between the Lab based INR value and Device based INR value at Week 12 or at Week 24.

A number of additional exploratory analyses taking into consideration the cross tabulations proposed by EMA were undertaken. The goal was to better understand the INR results obtained with the Alere INRatio device in the ROCKET AF trial, their correlation with the laboratory based INRs performed at weeks 12 and 24 as part of a pharmacodynamics study, and the impact of these correlations on the key efficacy and safety outcomes of the ROCKET AF trial. These additional exploratory analyses suggest that observed differences in warfarin event rates cannot be attributed to the adjustments made in warfarin dose based on discrepancy-related device INRs.

Sensitivity Analysis based on the FDA agreed Alere recall notice:

- Three sensitivity analyses, which were conducted based on the overall ROCKET AF study data by excluding subsets of the study population listed in the FDA agreed Alere recall notice, support the original efficacy and safety analyses of the ROCKET AF trial and confirm the positive benefit/risk profile of rivaroxaban). The analyses provide the most appropriate assessment of the impact of the Alere recall on the results of the ROCKET AF trial.

Device performance analysis in a subset of patients at Week 12 and Week 24:

1. In the ROCKET AF trial there were more than 160,000 INR-measurements (after Week 4) for subjects taking warfarin. The mean treatment duration was 580 days in the Warfarin group with the majority of subjects receiving treatment for at least 18 months.

   a) Only for 2 time points Lab based INR values are available with a corresponding Device based INR value (5766 values for Week 12 and 5507 for Week 24, Safety Analysis Set). 5048 patients had values obtained at both Week 12 and Week 24.

2. For this subset of data (i.e. Week 12 and Week 24 data) comparisons of the Lab based INR vs Device based INR were performed and it was revealed that there was a linear correlation for the Lab based INR value to the Device based INR value at Week 12 and 24 for the subjects With or Without any recall-related time of observation. Using discrepancy specifications adapted from

those defined in 2007 ISO 17593, nearly 90% of the Warfarin treated subjects in the ROCKET AF trial showed no discrepancy between the Lab based INR value and Device based INR value at Week 12 or at Week 24.

3. Additionally for these two dedicated time points, Bland-Altman plots for absolute and relative differences between the Lab based INR and Device based INR were prepared. With respect to the ratio of the difference over the Lab based INR median and (arithmetic) mean are of similar magnitude. Both show that on average the Device INR is about 13% lower than the Lab based INR.

4. In terms of differences for measurements With or Without recall related time, the data are consistent with previously provided information (please refer to attachment 1 for details) that nearly 90% of the Warfarin treated subjects in both groups had no discrepancy observed between the Lab based INR value and Device based INR value at Week 12 or at Week 24 (87% of INR fall in range).

5. Within the subgroups defined by the various Consistency definitions, the efficacy and safety event rates do not always reflect the expected effects on efficacy and safety outcome. Higher bleeding rates for Warfarin subgroups Without Consistency relative to the subgroup With Consistency are not matched by lower efficacy rates as one would hypothesize if INR would be driving the observed event rates.

6. With the additional exploratory analyses it was unexpectedly revealed that inconsistencies in INR measurements were associated with differences in rivaroxaban event rates as well as warfarin event rates, which also suggests that observed differences in warfarin event rates cannot be attributed to the adjustments made in warfarin dose based on discrepancy-related device INRs.

7. Further separation of this subset of data (Week 12 and Week 24) will not provide additional relevant information on the benefit/risk profile of rivaroxaban.

The MAH would like to reiterate that the initial response dated 15 October 2015)with the three sensitivity analyses constitutes the most sound assessment of the effect of the Alere correction notice on the ROCKET AF trial and provides the evidence that the results and conclusions from the ROCKET AF trial remain valid. The current further exploration (i.e. Bland-Altman plots, Cross Tabulation Lab INR vs Device based INR, and Concordance subgroup analyses based on Cross Tabulation), provided later in the procedure upon CHMP request, while related to the POC device, are not related to the Alere correction notice, have no clear relation to the overall ROCKET AF results, and have a number of limitations as outlined response to request 5.

2.4. Other questions raised by the CHMP

2.4.1. Request 3: Clarification whether the recalled device was used in other trials performed with Xarelto or any other development programmes, and submitted or intended for submission.

Summary of the MAH response:

Response

The INRatio PT/INR Monitor System was chosen to monitor INR in the ROCKET AF trial program. This trial program was the pivotal study program provided to EMA within the indication seeking variation procedure for the stroke prevention in non-valvular atrial fibrillation patients.

Bayer confirms that the INRatio and INRatio 2 PT/INR Monitor System was not chosen to monitor INR in any other ongoing or completed Bayer- or JRD-sponsored interventional or non-interventional rivaroxaban studies. Bayer cannot exclude that the INRatio and INRatio 2 PT/INR Monitor System was used by individual patients or investigators to monitor INR in open label rivaroxaban studies like XALIA or EINSTEIN PE / DVT.

Bayer further confirms that the INRatio and INRatio 2 PT/INR Monitor System was not chosen to monitor INR in any other development programmes submitted or intended for submission to regulatory agencies by Bayer. Bayer has not purchased the INRatio and INRatio 2 PT/INR Monitor System for any other clinical study.

The CHMP considered that the requested confirmation has been provided and considered the issue closed.

2.4.2. Request 4: Information whether other Regulatory authorities have been informed and subsequent actions were taken in any territory.

With a cut-off date of October 11, 2015, the following 26 countries / Health Authorities were informed on a global basis (Europe/EMA, Sweden (rapporteur), Germany (co-rapporteur), US/FDA, Switzerland, Canada, Australia, Japan, New Zealand, China, Israel, Macedonia, Serbia, Montenegro, Kosovo, Singapore, Korea, Taiwan, Malaysia, France, UK, Norway, Denmark, Iceland, Ireland, Netherlands) per local regulations and practices.

On October 01, 2015 Bayer received a written request by EMA in co-operation with the rapporteur and the co-rapporteurs to provide further information as outlined in this response document. On October 09, 2015 a request letter for further information by the MHLW (Ministry of Health, Labour, and Welfare, Japan) was received as well. So far, no other subsequent actions have been taken in other jurisdictions (cut-off date: October 11, 2015).

The CHMP noted the information that authorities in the concerned countries have recently been informed by the MAH on the issue.

3. Overall conclusion and impact on the benefit/risk balance

The manufacturer of the point of care INR device (Alere Inc.TM) has identified that the device in certain clinical conditions gave inappropriately low INR values. The device was used for titration of the warfarin doses in the rocket AF study. The MAH of Xarelto was requested to provide clarification and further information in order to assess the potential impact on the pivotal Rocket AF study results supporting the the indication for rivaroxaban in the prevention of thromboembolism in non-valvular atrial fibrillation (The "Rocket AF" study). All requested analyses have been provided.

The inappropriately low INR values observed with the device as indicated in the device correction notice of Alere were in some instances reported to be clinically relevant leading to increases of anti-vitamin K agents with resulting increased bleeding tendency. Thus it was considered relevant to further investigate the impact of potentially inaccurate device INR estimations ,that could have had on the study results. A major concern would be that the described deficiencies could lead to an increased bleeding rate in the control arm of the trial (warfarin) which would potentially hamper the interpretations of the differences between treatment arms. The investigations performed and possible conclusions that can be made from them are summarised below

Numerous extensive sensitivity analyses have been performed in order to estimate the potential impact on the study outcome.

The MAH initially performed three sensitivity analyses were based on the information provided by the INR device manufacturer (Alere) and taking both chronic and acute clinical conditions into account. Patients who were judged to have had any recall related clinical condition ahead of any endpoint event are excluded in the second of these sensitivity analyses which resulted in an exclusion of approximately 31% of the patients. The first and third sensitivity analyses are focussing on the time periods when such events occurred and fewer patients are excluded in these analyses. In all three analyses minor changes in the hazard ratios between the rivaroxaban and warfarin study arms were observed as compared to the original analyses. For the sensitivity Analysis 2 considered to be the most conservative analysis of the three analyses provided, warfarin was almost nominally statistically significantly better than rivaroxaban for the principal safety endpoint, major or clinically relevant bleedings (p = 0.092, HR: 1.08 (CI: 0.99, 1.17)).

As can be seen in Table 2, the pattern of bleeding components was different between rivaroxaban and warfarin with death due to bleeding, critical organ bleeding and intracranial haemorrhage being nominally in favor of rivaroxaban. This differential pattern remained unchanged in the sensitivity analyses cited above. (table 8). There was no indication that the most serious bleeding events on warfarin were related to recall related clinical conditions or events.

There are uncertainties related to the sensitivity analyses above discussed. In particular the available information is probably somewhat limited e.g. fibrinogen levels have not been measured regularly and is rarely measured in clinical routine. The analyses were based on the assumption that INR values are only affected in case of recall related conditions. This issue was addressed by additional analyses.

In a second step, these additional sensitivity analyses were provided upon request from CHMP. These analyses were focussing on the paired samples that were taken simultaneously at weeks 12 and 24 of the trial. In addition to the Alere device estimated INR, a laboratory prothrombin time measurement was made. The primary intention of this exercise was to evaluate the potential impact of rivaroxaban on prothrombin time during the Rocket AF trial. In retrospect, however, in view of the INR device issue under discussion, an INR ratio could be calculated allowing comparisons of the INR values as estimated by the two methods. (INR device and INR laboratory values calculated from the two sampling time points at week 12 and week 24).

The MAH has, as requested, provided Bland-Altman plots comparing the estimations made by the two methods (INR device and INR calculation based on the laboratory values). From visual impression of the Bland-Altman plots, it indicates that the majority of discrepancies were related to lower device based INR value as compared to the Lab INR value. This was also further elucidated when looking at the cross tabulations made based on clinically relevant INR categories which showed that 34% (1961 out of 5766) of the measurements has a lower device INR compared with lab INR (discordant values).The reverse, that the device INR was higher compared with the Lab INR, was observed in only 4% of the measurements (discordant values). However, consistency,(i.e. that the measurements fell in the same INR category), was observed in 62% of the instances (concordance), The subjects with and without concordance whether further investigated to explore the impact of the device values. The principal safety endpoint event rates for subjects with and without concordance at Week 12 and at Week 24 were dissimilar (12.9/100 pt-yrs vs 14.0/100 pts-yrs, and 12.6/100 pt-yrs vs 13.6/100 pts-yrs, respectively), although closer with and without concordance at both week 12 & 24 (12.2/100 pt-yrs vs 12.9/100 pts-yrs, respectively). Based on these analyses modified event rates in the warfarin arm were calculated and compared with the observed event rates in the rivaroxaban arm. The impact on the originally reported Hazard Rate Ratios was small as seen in table 8.

Theoretically, the differences observed could, on group level, depend on higher warfarin doses than necessarily needed in patients with discrepancy. However, the differences above discussed are rather

small, and when taking also efficacy data in these analyses into account it appears reasonable to conclude that the characteristics and risks in these non-randomised groups differ from the overall study population. This is further supported by the additional analyses provided by the MAH where a very similar pattern for the safety end-points is observed in the rivaroxaban patients and where doses were not adjusted. Furthermore, when looking at the patients who have no paired samples it becomes even more apparent that the sensitivity analyses reflect subsets of patients that are not representative for the over-all study population.

The CHMP acknowledges that there are also limitations related to the above analyses as they only represent two specific time points during treatment approximately 11.000 out of 160.000 INR measurements which were performed during the study.

A single lab INR cannot be regarded as a true "golden standard". However, there is since long an awareness of the important influence different thromboplastins with different sensitivity have on the size of the prothrombin time prolongation. The INR calculations compensate for these differences. Among the frequently used routine coagulation tests laboratory INR testing is probably one of the best standardised methodologies. Point of care devices for INR estimations are generally speaking probably somewhat less well standardized however, they have important feasibility advantages together with sufficient reliability which could justify their use preferable in clinical studies or in clinical routine for individual patients. There is also available extensive external evidence, important to take into consideration when evaluating the bleeding events in the Rocket AF trials. The incidences and the characteristics of bleedings induced by rivaroxaban in comparison with warfarin were characterised in other large clinical trials supportive of the indication for secondary prevention of DVT/PE. The Alere INR device was not used in these studies. When looking at the characteristics of the study populations in the DVT/PE prevention studies and the doses of rivaroxaban used , they are rather similar to the Rocket AF trial. The targeted INR was also identical. Therefore, the DVT/PE prevention studies are considered relevant for external validation purposes. It can be noted that in the pooled analyses of the DVT and PE studies, the major and clinically relevant bleedings were numerically fewer and major bleedings significantly fewer in the rivaroxaban arm (HR 0.93, 95% CI 0.81;1.06 and HR 0.55, CI 0.34, 0,79, respectively). The bleeding characteristics were similar to those observed in the Rocket AF trial.

The bleeding rates and the bleeding characteristics in the warfarin arm in the Rocket trial is also within the expected range as the one described in the literature. Furthermore the analyses of the components of the primary safety endpoint, critical organ bleeding, fatal bleeding, ICH, clearly favoured the rivaroxaban group as did all-cause mortality. These differences remained largely unaffected in the sensitivity analyses performed and taking also the size of the differences into account they are considered to be robust. They are also consistent with the bleeding pattern seen for the other available oral Xa-inhibitors.

The conclusions derived from the Rocket AF trial are supported by the performed sensitivity analyses and by external evidence from large trials in other therapeutic areas comparing the efficacy and safety of rivaroxaban vs warfarin. The incidence of major and clinically relevant bleedings among the warfarin treated patients in the Rocket AF trial compares also well with the bleeding incidences observed in recently performed trials in AF populations. Incidences of major bleedings in warfarin arms in the dabigatran, apixaban and edoxaban pivotal trials in atrial fibrillation were 3.6, 3.1 and 3.4%, respectively, as compared to 3.5% in the Rocket AF trial.

It is agreed with the MAH that further analyses are not expected to provide additional information of substantial value.

In conclusion, the CHMP considered that there is sufficient evidence to conclude that the benefit/risk balance remains unchanged and favourable for treatment with rivaroxaban in the prevention of thromboembolism in non-valvular atrial fibrillation.

The CHMP also considered that the information provided in the SmPC is currently appropriate and does not warrant any amendment. The modified calculated INR values discussed in this report should be viewed as informative for the assessment of the INR values provided by the faulty device. Taking into account the extent of the potential impact of the modified INR values and the limitations of the post hoc sensitivity analyses, it is considered that an update in the SmPC section 5.1 is not relevant and would provide more confusion to the prescriber.