Paper DH10

Managing Privacy Risk in Data and Document Sharing: Lessons Learned Balancing Progress & Perfection

d-wise™
Clinical Data and Document sharing is no longer ascending the agenda, it is on the agenda. As sponsors move forward, protecting patient privacy is fundamental and non-negotiable. Sharing data is fundamentally about disassociating the patient from the data by generalizing or masking the content in datasets and documents. Techniques for masking or generalizing PII are beginning to stabilize, simultaneously the matrix of concerns broadens. Sponsors require approaches to measure risk that are quantitative, repeatable, and scalable. There is no one-size-fits-all approach, but – as the saying goes – perfect is the enemy of progress. How can sponsors balance utility and privacy while meeting various global requirements? How should sponsors manage data they share considering what’s already been shared? What techniques exist to support sponsors in navigating the reality of human error and the limits of technology? This talk will feature lessons learned in pursuit of progress in clinical document and data sharing.

INTRODUCTION & BACKGROUND

The Data Transparency landscape comes with a myriad of requirements globally with more than 90 registries around the world [1]. It has also greatly evolved over the past few years with respect to sharing Clinical Study Reports (CSR), starting with Regulators where we’ve seen in addition to PMDA (2013, [2]) and EMA (2014 [3]), Health Canada (2019 [4]) and FDA (2018 [5]) defining their own initiatives (Not effective yet at the time this paper is written).
The table below highlights the differences in scope, processes and technical requirements across Agencies’ requirements when sharing CSRs as part of the submissions process.

<table>
<thead>
<tr>
<th>Submission Driven Clinical Data Publication</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EMA Policy 0070</strong></td>
</tr>
<tr>
<td>Who does the PII and CCI work?</td>
</tr>
<tr>
<td>Who has final say on PII?</td>
</tr>
<tr>
<td>Who has final say on CCI?</td>
</tr>
<tr>
<td>Technical Requirements</td>
</tr>
<tr>
<td>Individual Patient Level Details</td>
</tr>
<tr>
<td>Which modules are made public</td>
</tr>
</tbody>
</table>

Multiple scenarios for sharing are now commonplace:

- **Regulatory Sharing** - The main objective of Data Sharing for Regulators is to provide to the general public, academics and other stakeholders access to data that was reviewed by Agencies during the submission process. The informational or research value of the post-anonymized data (the “Data Utility”) is differs based on the level of de-identification applied – some approaches preserve Data Utility, others reduce Data Utility, and either may be appropriate in a given disclosure scenario based on privacy and risk considerations. Agencies (EMA and Health Canada) also ask sponsors to justify how they optimized Data Utility after de-identification and document it in the Anonymization Report.

- **Research Sharing** - Aside from regulatory drivers, industry is more and more choosing to participate in sharing voluntarily. Since 2013, a number of sponsors started to voluntarily share their data with academia for secondary-purpose research. EFPIA and PhRMA also released the “Principles for Responsible Clinical Trial Data Sharing” [6] which has been endorsed by many sponsors. The sponsors following these principles commit to sharing upon request from qualified scientific and medical researchers patient-level clinical trial data, study-level clinical trial data, and protocols from clinical trials in patients for medicines and indications approved in the United States (US) and the European Union (EU) as necessary for conducting legitimate research.

- **Publication Requirements** - As of 1 July 2018, manuscripts submitted to ICMJE journals [7] that report the results of clinical trials must contain a Data Sharing statement. Such Data Sharing statements must indicate in particular whether individual deidentified participant data will be shared and whether additional, related documents will be available (e.g., study protocol, statistical analysis plan, etc.).
This paper discusses lessons learned at AstraZeneca addressing Data Transparency requirements where balancing Patients Privacy and Data Utility is at the heart of the process and explores certain features of the Blur product to support Data Anonymization activities.

QUALITATIVE VS. QUANTITATIVE RISK ASSESSMENT

EMA refers in their external guidance ([8] Chapter 2, section 5.1 Data Utility) to a qualitative approach and use of redaction in an initial phase until sponsors gain more experience and can use quantitative approach and anonymization. The guidance suggests that there is a period where sponsors can develop their skills and approach to transition to quantitative and anonymization techniques. Health Canada Guidance [9] also encourage sponsors to use a quantitative risk analysis over a qualitative one. Whereas said agencies show favor for quantitative approaches, both FDA and PMDA seems to use and refer respectively to a qualitative risk assessment.

Qualitative risk assessments often lead to redactions in documents released in the public domain [10] while for Individual Patient Data, in the absence of risk quantification compared against a pre-defined threshold, data can only be claimed to be pseudomized rather than anonymized. A quantitative risk assessment both enables anonymization of Individual Patient Data in both documents and data.

QUALITATIVE RISK ASSESSMENT

The table below provides an example of how a qualitative risk assessment can be conducted considering “How Identifying” and “How Sensitive” data is. Based on the risk level (Low, Medium, High), redaction may apply to more or less quasi-identifiers. Both TransCelerate [11] and PhUSE [12] have tried to describe how to approach a qualitative assessment and provide further guidance.
QUANTITATIVE RISK ASSESSMENT

In the case of public disclosure, a Quantitative Risk Assessment measures (numerically) the maximum risk of re-identification across all patients and is typically consider relative to a pre-defined target risk tolerance threshold. EMA Policy 0070 guidance [8] advises to use a default threshold of 0.09 and it is possible to select other threshold with a justification. Measuring reidentification risk for a patient within population can be accomplished by appropriately segregating the population into equivalence classes. To attain the targeted threshold risk of 0.09 (or less) means that any patient is in an equivalence class of at least 11 for the selected quasi-identifiers (or in other words that any individual should not stand alone in a group of less than 11 individuals with the same characteristics as them).

The risk calculation relies on both the selection of quasi-identifiers and the selection of a reference population ([9], [13], [14]). Typical considerations in determining the reference population include (but are not limited to): 1) the study population, 2) the sponsor’s similar studies population, 3) the similar studies population and 3) the disease or 4) geographical population.

RISK GIVEN THE CONTEXT OF DISCLOSURE AND NATURE OF ADVERSARY

In the case of public disclosure such as data shared under EMA Policy 0070, the maximum risk metric is advised to be used and compared to a threshold (e.g. 0.09). This means that all patients must be in equivalence classes of at least 11 people with the same characteristics. Under other contexts such as when data is shared with researchers through a portal and with a data sharing agreement, the average risk metrics can be used where the average risk across all patients is compared to the threshold ([13], [14]). In addition, technical and contractual controls lower the probability of the different types of attacks (see Figure below) while the probably of attack for public disclosure is set to 1.
For all \( i \), \( P(\text{ReID} \cap \text{Attack}) = P(\text{ReID}/\text{Attack}) \times P(\text{Attack}) \leq \text{Threshold} \)

<table>
<thead>
<tr>
<th>Attack</th>
<th>Example</th>
<th>Factors influencing ( P(\text{Attack } i) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Attempt</td>
<td>Researcher attempts at re-identifying patients</td>
<td>Mitigating Controls</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Motives &amp; Capacity</td>
</tr>
<tr>
<td>2. Acquaintance</td>
<td>Researcher spontaneously recognizes patients</td>
<td>Study Patients Prevalence</td>
</tr>
<tr>
<td>3. Breach</td>
<td>A rogue organization “hacks” in the portal and retrieve the data</td>
<td>Security Practice at Data Recipient</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Portal Security)</td>
</tr>
</tbody>
</table>

- \( P(\text{ReID}/\text{Attack}) \) is controlled through data de-identification
- \( P(\text{Attack}) \) is dependant on disclosure context
  - \( P(\text{Breach}) = 1 \) for public disclosure

**UTILITY THROUGH ANONYMIZATION**

Whereas Data Utility is nonexistent when content is redacted, anonymized content allows Data Utility to be preserved. Quantitative risk assessment is a technique that allows risk aware anonymization as an informed, deliberate alternative to pure redaction. Quantitative risk-based approaches:

- should lead in principle to less redaction in favor of further use of anonymization techniques (generalization, offsetting, etc.),
- should result in an increase in the Data Utility after anonymization, and afford the data controller measurable privacy control have over the disclosed data.

The figure below also highlights the competing nature of addressing both Data Utility and risk where ability to quantifying the risk and Data Utility enables to find the “sweet spot” between the two, which is illustrated in the next section.
QUANTIFYING DATA UTILITY

Data Utility is subjective to the intended usage (simply, Data Utility is in the eye of the beholder). In the case of a disclosure following a research request, it is possible to evaluate which variables are important to preserve and possibly rank them. In the case of a public disclosure, there could be a myriad of Data Recipients with legitimate purpose and it is not possible to predict and address all needs that could be competing. A Data Utility measure referred to as Precision [15] can be used to quantify the Data Utility according to how much/little distortion has been applied to the data. The two tables below describe how the Precision is computed considering a single variable or multiple ones.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Rule / Damage</th>
<th>Precision / Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Keep unchanged</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>5 Year age bands</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>10 Year age bands</td>
<td>0.50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Rule</th>
<th>Precision</th>
<th>Variable</th>
<th>Rule</th>
<th>Precision</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Keep</td>
<td>1.00</td>
<td>Sex</td>
<td>Keep</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>5 Year age band</td>
<td>0.75</td>
<td>Keep</td>
<td>1.00</td>
<td>0.875</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 Year age band</td>
<td>0.50</td>
<td>Keep</td>
<td>1.00</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adult Child Age Band</td>
<td>0.25</td>
<td>Keep</td>
<td>1.00</td>
<td>0.625</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Drop</td>
<td>0.00</td>
<td>Keep</td>
<td>1.00</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Keep</td>
<td>1.00</td>
<td>Drop</td>
<td>0.00</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 Year age band</td>
<td>0.75</td>
<td>Drop</td>
<td>0.00</td>
<td>0.375</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 Year age band</td>
<td>0.50</td>
<td>Drop</td>
<td>0.00</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adult Child Age Band</td>
<td>0.25</td>
<td>Drop</td>
<td>0.00</td>
<td>0.125</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Drop</td>
<td>0.00</td>
<td>Drop</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>
In addition, it is possible to introduce and assign weights to the different quasi-identifiers as reflected in table below in order to reflect the relative importance of each of them. This may be driven by the need expressed in a research request or in the case of public disclosure (e.g. EMA Policy 0070), these weights can be assigned considering which quasi-identifiers are confounding/prognostic or stratification variables and should be prioritized over others that are not.

In table below, a weight of 2 is assigned to Sex while Age gets a weight of 1. When looking at the overall Precision, we can see that any deterioration of Sex (here “Drop”) leads to a large reduction of the overall Precision; Maximum Precision is 0.333 when Age remains unchanged (“Keep”).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Rule</th>
<th>Precision</th>
<th>Variable</th>
<th>Rule</th>
<th>Precision</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (1)</td>
<td>Keep</td>
<td>1.00</td>
<td>Sex (2)</td>
<td>Keep</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>5 Year age band</td>
<td>Keep</td>
<td>0.75</td>
<td>Keep</td>
<td>1.00</td>
<td>0.91666</td>
<td></td>
</tr>
<tr>
<td>10 Year age band</td>
<td>Keep</td>
<td>0.50</td>
<td>Keep</td>
<td>1.00</td>
<td>0.833</td>
<td></td>
</tr>
<tr>
<td>Adult Child Age Band</td>
<td>Keep</td>
<td>0.25</td>
<td>Keep</td>
<td>1.00</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>Drop</td>
<td>Keep</td>
<td>0.00</td>
<td>Drop</td>
<td>0.00</td>
<td>0.667</td>
<td></td>
</tr>
<tr>
<td>Keep</td>
<td>1.00</td>
<td></td>
<td>Drop</td>
<td>0.00</td>
<td>0.333</td>
<td></td>
</tr>
<tr>
<td>5 Year age band</td>
<td>0.75</td>
<td></td>
<td>Drop</td>
<td>0.00</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>10 Year age band</td>
<td>0.50</td>
<td></td>
<td>Drop</td>
<td>0.00</td>
<td>0.1667</td>
<td></td>
</tr>
<tr>
<td>Adult Child Age Band</td>
<td>0.25</td>
<td></td>
<td>Drop</td>
<td>0.00</td>
<td>0.0833</td>
<td></td>
</tr>
</tbody>
</table>

The Precision is a simple metric that enables comparison of different set of rules from a Data Utility perspective across multiple quasi-identifiers.

The idealized outcome is when Data Utility as measured by Precision is optimized subject to minimizing risk given risk must be at or below the threshold. In practice, the de-identifier would first evaluate which sets of rules leads to a maximum risk under the threshold (e.g. 0.09) and then among the set of rules that qualify from a risk stand-point, compare their Data Utility using the Precision metrics and select the final set of rules to be used to anonymize the data. This is illustrated in next section.
This section illustrates how to implement the concepts of risk and utility quantification using features from the Blur product (a commercial anonymization software). The methodology follows 2 main steps:

1. Find all combinations of sets of rules (anonymization techniques) that meet the risk threshold
2. From the set that meet the risk threshold, select the set of rules leading to the highest Precision

There are other preparation steps such as defining the hierarchies for each variable [15] and conducting an assessment of the disclosure context [14] to set the probability of attack as well as final steps consisting of applying the selected set of rules to the dataset and/or document that are not described here. The section focuses on the heart of the anonymization process where risk and utility are measured and balanced against each other.

**STEP 1: FIND ALL SET OF RULES THAT MEET THE RISK THRESHOLD**

After the user has defined the possible rules to apply to each IPD-relevant variable (also called hierarchies) at stake in the risk quantification, the Blur Simulation Module enables users then:

1. to simulate all possible combination of rules across variables
2. to visualize the different set of rules and their respective re-identification risk (Column PoSA – Probability of Successful Attack) and respective Precision for the dataset.

In the image below, the measured risk for a given set of rules decreases from top to bottom (see the column titled “PoSA” meaning probability of successful attack). The sets of rules with a re-identification risk exceeding the threshold (0.09 in the example below) and are highlighted in red. The user focuses on the sets of rules with achieve the target threshold (in grey at the top, outlined by the green bounding box).
Note that the % of patients exceeding the threshold (column “% PoSA Exceeded”) displayed on the right-hand side can help adjusting the rules in subsequent iterations should an initial set of rules with higher Data Utility be just above the threshold.

**STEP 2: SELECT THE SET OF RULES LEADING TO THE HIGHEST PRECISION**

The user then selects the set of rules with the highest Precision among the sets of rules that meet the 0.09 threshold. Figure below also illustrates the different weights that were assigned for each quasi-identifier to compute the Precision.
This section explores sample CSRs from AstraZeneca's Policy 0070 submission for Tagrisso that were published in 2017. At the time of publication, a qualitative approach coupled with redaction was used. The select CSRs have been anonymized again using the Blur product and a quantitative approach and compared to the originally reports published in 2017.

The identification and anonymization of direct and quasi identifiers in documents is supported by Natural Language Processing [16] which enables anonymization rules to be applied to text within unstructured artifacts such as CSRs (or other documents). Such transformations would not be possible without the use of such technology unless certain sections of the document are re-written or re-generated [12].

After following the process described in the previous section where different set of rules are evaluated for risk of re-identification and Data Utility through the Precision metric, a set of rules is selected for each given study.

Examples from select sections are available in Appendices 1 where sections redacted and published on the EMA portal [17] are compared to same sections anonymized using the Blur product (not published). The appendix shows a comparison and help illustrate the following:

- How recoding of subject IDs instead of redaction enables to follow a patient across different sections of documents. Note that all these sections are also enhanced too for other patient’s data compared to when only qualitative assessment and redaction methods were used.
- Release anonymized in-text "mini-narratives" and narratives where subject ids are recoded, dates offset, gender is retained, and age generalized rather than fully redacted.
- Release in-text events listings where subject ids are recoded, dates offset, gender is retained, and age generalized rather than being partially redacted.
Recoding Subject IDs and offsetting dates enables the story around patients to be built and better understood [18] across sections whether they are body text, narratives or in-text events listings while retaining or generalizing other demographics (e.g. gender and age) unlock another level of clinical data granularity. As an example, offset dates preserve elapsed days between events without revealing the genuine date of actual event whereas redacted dates completely eliminate all informational value regarding the timing of events.

It must be noted as well that certain sections such as narratives with the enhanced control that a quantitative assessment provides are now released with some transformations rather than being fully redacted.

CONCLUSION

In order to increase Data Utility and support further use of anonymized data, moving from redaction to anonymization is key while ability to measure and quantify the risk increase privacy controls over a data disclosure. Solutions both supporting quantification of risk and Data Utility coupled with Natural Language Processing ([12], [16]) capability can help addressing such needs for both documents and IPD datasets.

Anonymization techniques such as recoding, date offsetting or generalization of age or geographical location enable data recipients to follow patients within and across documents, retain sequence and distances between events and interventions and release richer data. Such anonymization techniques are both recommended to be used under EMA Policy 0070 guidance [8] and Health Canada Public Disclosure of Clinical Documents guidance [9].

In this ever-changing Data Transparency landscape [19], sponsors should be able to create the anonymized package once and submit to multiple agencies assuming the different technical requirements allow this. Having multiple standards driven by different requirements from Agencies could pose challenges for companies who may need to conduct similar efforts more than once. The need for automation and scalable solutions becomes then even more critical.
REFERENCES


[3] EMA Policy 0070 of 2nd October 2014:

http://gazette.gc.ca/rp-pr/p1/2017/2017-12-09/html/reg3-eng.html

[5] FDA Clinical Data Summary Program page:
https://www.fda.gov/drugs/developmentapprovalprocess/ucm589210.htm


[8] EMA Policy 0070 External Guidance versions:


[10] Analysis of CSRs already published, PhUSE Review – Lukasz Kniola, November 2017


https://www.phusewiki.org/docs/Conference%202017%20TT%20Papers/Paper%20TT07%202017%20NEW.pdf

[17] EMA portal for the publication of clinical data
https://clinicaldata.ema.europa.eu/

[18] “EMA Policy 0070: Data Utility in Anonymised Clinical Study Reports (CSRs)”, Ferran, Nevitt – DH04, PhUSE 2017 Annual Conference
ABOUT THE AUTHORS

Julie Holtzople is the Clinical Trial Transparency Operations Director at AstraZeneca. Julie is responsible for developing and implementing the processes, standards, systems and tools to ensure all AstraZeneca Clinical Trial Transparency policies and standards are fully realized. Currently, Julie is spending a lot of time working on the implementation of Anonymization of Clinical Documents and the delivery of Lay Language Summaries. Prior to working at AstraZeneca, Julie spent almost a decade working as a consultant for Ernst & Young and Booz Allen Hamilton. Julie is a graduate of Virginia Tech.

Stephen Baker is Director of Client Management for d-wise. His focus is providing advisory and leadership services to Life Sciences clients to enhance the business architecture that supports clinical research.

The opinions in this paper are our own.

ACKNOWLEDGEMENTS

We would like to thank Jean-Marc Ferran (Qualiance) for his help writing the paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:
Julie Holtzople
AstraZeneca
Email: julie.holtzople@astrazeneca.com

Stephen Baker
d-wise Technologies
Email: Stephen.Baker@d-wise.com

Brand and product names are trademarks of their respective companies.
Clinical Study Report  
Drug Substance AZD9291  
Study Code D160000000  
Edition Number 1  
Date 17 August 2015  

Table 10  
Important protocol deviations – key details

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Protocol deviation coded term</th>
<th>Included in PK analysis (Y/N)</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPD</td>
<td>Study procedures criteria</td>
<td>Yes</td>
<td>Site did not record meal end time on Day 1. Therefore, unable to determine whether meal was consumed in the 30-minute window. The patient consumed 100% of the meal.</td>
</tr>
<tr>
<td></td>
<td>Study procedures criteria</td>
<td>Yes</td>
<td>Site did not record meal end time on Day 10. Therefore, unable to determine whether meal was consumed within 30 minutes. The patient consumed 100% of the meal.</td>
</tr>
<tr>
<td></td>
<td>Concomitant medication</td>
<td>Yes</td>
<td>Patient excluded from PK analysis for Period 2 as they started CYP3A4 inhibitor (amlodipine) on Day 9 (see Section 6.5.1).</td>
</tr>
<tr>
<td></td>
<td>Eligibility criteria</td>
<td>No</td>
<td>The patient was incorrectly enrolled as had cardiomyopathy at baseline.</td>
</tr>
<tr>
<td></td>
<td>Study procedures criteria</td>
<td>No</td>
<td>Patient was only able to consume 50% of the meal on Day 10 which is less than the protocolled minimum of 75%. This patient was excluded from the PK analysis set (see Section 6.3.2).</td>
</tr>
<tr>
<td></td>
<td>Concomitant medication</td>
<td>Yes</td>
<td>Patient excluded from PK analysis for Period 2 as they stopped CYP3A4 inhibitor (amlodipine) on Day 12 (last dose Day 11) (see Section 6.5.1).</td>
</tr>
<tr>
<td></td>
<td>IP compliance</td>
<td>Yes</td>
<td>Patient was dosed 15 minutes after the start of the meal on Day 10. The protocol specified AZD9291 dose should be administered 30 minutes after the start of the meal consumption. The PK data were included in the analysis as the patient had consumed all of the food by the time of dosing so it was considered that there was enough food in the stomach to assess the interaction with food.</td>
</tr>
<tr>
<td></td>
<td>IP compliance</td>
<td>Yes</td>
<td>Patient was dosed 10 minutes after the start of the meal on Day 10. The protocol specified time is 30 minutes. The PK data were included in the analysis as the patient had consumed all of the food by the time of dosing so it was considered that there was enough food in the stomach to assess the interaction with food.</td>
</tr>
</tbody>
</table>

Source: Appendices 12.2.2.1, 12.2.3.1, 12.2.4.2, 12.2.4.12, and 12.2.7.1  
Abbreviations: CYP = cytochrome P450, IP = investigational product, PK = pharmacokinetics.
### Table 10

**Important protocol deviations – key details**

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Protocol deviation coded term</th>
<th>Included in PK analysis (Y/N)</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>SX2H5W7</td>
<td>Study procedures criteria</td>
<td>Yes</td>
<td>Site did not record meal end time on Day 1. Therefore, unable to determine whether meal was consumed in the 30-minute window. The patient consumed 100% of the meal.</td>
</tr>
<tr>
<td>WI7PO4M3</td>
<td>Study procedures criteria</td>
<td>Yes</td>
<td>Site did not record meal end time on Day 10. Therefore, unable to determine whether meal was consumed within 30 minutes. The patient consumed 100% of the meal.</td>
</tr>
<tr>
<td>NOEATO4L</td>
<td>Concomitant medication</td>
<td>Yes</td>
<td>Patient excluded from PK analysis for Period 2 as they started CYP3A4 inhibitor (amiodipine) on Day 9 (see Section 6.5.1).</td>
</tr>
<tr>
<td>TBBB8KD</td>
<td>Eligibility criteria</td>
<td>No</td>
<td>The patient was incorrectly enrolled as had cardiomyopathy at baseline.</td>
</tr>
<tr>
<td>LOZ5XDL</td>
<td>Study procedures criteria</td>
<td>No</td>
<td>Patient was only able to consume 50% of the meal on Day 10 which is less than the protocol defined minimum of 75%. This patient was excluded from the PK analysis set (see Section 6.3.3).</td>
</tr>
<tr>
<td>RSCUFZQ</td>
<td>Concomitant medication</td>
<td>Yes</td>
<td>Patient excluded from PK analysis for period 2 as they stopped CYP1A4 inhibitor (amiodipine) on Day 12 (last dose Day 11) (see Section 6.5.1).</td>
</tr>
<tr>
<td>361L21AL</td>
<td>IP compliance</td>
<td>Yes</td>
<td>Patient was dosed 15 minutes after the start of the meal on Day 10. The protocol specified AZD9291 dose should be administered 30 minutes after the start of the meal consumption. The PK data were included in the analysis as the patient had consumed all of the food by the time of dosing so it was considered that there was enough food in the stomach to assess the interaction with food.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yes</td>
<td>Patient was dosed 10 minutes after the start of the meal on Day 10. The protocol specified time is 30 minutes. The PK data were included in the analysis as the patient had consumed all of the food by the time of dosing so it was considered that there was enough food in the stomach to assess the interaction with food.</td>
</tr>
</tbody>
</table>

Source: Appendices 12.2.2.1, 12.2.3.1, 12.2.3.2, 12.2.4.12, and 12.2.7.1

Abbreviations: CYP = cytochrome P450, IP = investigational product, PK = pharmacokinetics.
6.3 Patients analysed (analysis sets)

The analysis sets and the number of patients in each analysis set are summarised in Table 11.1.3. Definitions of the analysis sets are given in Section 5.7.2.

6.3.1 Safety analysis set

The safety analysis set included all 38 patients who received at least one dose of AZD9291.

6.3.2 Pharmacokinetic analysis set

The PK analysis set included 34 of the 38 patients who received AZD9291 and had at least one quantifiable plasma concentration collected post-dose without any protocol deviations/events affecting the PK. Table 10 and Table 11 provide a breakdown of patients who are excluded from the PK analysis in 1 or both periods.

Four patients were excluded from the PK analysis set due to protocol deviations/events affecting both treatment periods:

- **PFD** The patient had an SAE of atrial flutter in Period 1, and was administered medications diltiazem and verapamil, which were CYP3A4 modulators (see Section 6.5.1). Both the SAE and the medications administered could have meaningfully affected the PK for this period. The patient was only able to consume 50% of the meal on Day 10 which is less than 75% required by the protocol (also reported as a protocol deviation, see Table 10).

- **PFD** Patient vomited within 6 hours after dosing in Period 1 and Period 2.

- **PFD** Patient excluded from PK analysis for Periods 1 and 2 as they started CYP3A4 inducer (methylprednisolone) 1 day before dose on Day 1. Since the CYP3A4 inducer was not given for long enough before the first dose, any potential induction may have had a different effect on the first treatment period compared against the second treatment period.

- **PFD** Patient excluded from PK analysis for Periods 1 and 2 as patient stopped CYP3A4 inducer (echinacea) 4 days before dose on Day 1. Since the CYP3A4 inducer was not stopped long enough for an adequate washout time before the first dose, any potential reversal of induction may have had a different effect on the first treatment period compared against the second treatment period.

Fasted treatment period

Of the 34 patients in the PK analysis set, 32 had evaluable records for the Fasted treatment. Two patients were excluded from the Fasted treatment, PFD started a CYP3A4 inhibitor [perindopril/amlopidine] during Period 2 [Day 9] and PFD vomited after dosing in Period 2 [since time of AE was not available, patient was excluded from the treatment period].
6.3 Patients analysed (analysis sets)

The analysis sets and the number of patients in each analysis set are summarised in Table 11.13. Definitions of the analysis sets are given in Section 5.7.2.

6.3.1 Safety analysis set

The safety analysis set included all 38 patients who received at least one dose of AZD9291.

6.3.2 Pharmacokinetic analysis set

The PK analysis set included 34 of the 38 patients who received AZD9291 and had at least one quantifiable plasma concentration collected post-dose without any protocol deviations/events affecting the PK. Table 10 and Table 11 provide a breakdown of patients who are excluded from the PK analysis in 1 or both periods.

Four patients were excluded from the PK analysis set due to protocol deviations/events affecting both treatment periods:

- **7BBBKKAD**: The patient had an SAE of atrial flutter in Period 1, and was administered medications diltiazem and verapamil, which were CYP3A4 modulators (see Section 6.5.1). Both the SAE and the medications administered could have meaningfully affected the PK for this period. The patient was only able to consume 50% of the meal on Day 10 which is less than 75% required by the protocol (also reported as a protocol deviation, see Table 10).

- **ZPKJJDYD**: Patient vomited within 6 hours after dosing in Period 1 and Period 2.

- **QR3D0HNS**: Patient excluded from PK analysis for Periods 1 and 2 as they started CYP3A4 inducer (methylprednisolone) 1 day before dose on Day 1. Since the CYP3A4 inducer was not given for long enough before the first dose, any potential induction may have had a different effect on the first treatment period compared against the second treatment period.

- **3KE00QWB**: Patient excluded from PK analysis for Periods 1 and 2 as patient stopped CYP3A4 inducer (echinacea) 4 days before dose on Day 1. Since the CYP3A4 inducer was not stopped long enough for an adequate washout time before the first dose, any potential reversal of induction may have had a different effect on the first treatment period compared against the second treatment period.

**Fasted treatment period**

Of the 34 patients in the PK analysis set, 32 had evaluable records for the Fasted treatment. Two patients were excluded from the Fasted treatment (**NOEATC4L** started a CYP3A4 inhibitor [perindopril/amlodipine] during Period 2 [Day 9] and **HYB00RM** vomited after dosing in Period 2 [since time of AE was not available, patient was excluded from the treatment period]).
6.5  Use of concomitant medication and treatment compliance

6.5.1  Concomitant medication after study entry

Disallowed and allowed concomitant medications administered during the study are summarised in Tables 11.1.15 and 11.1.16, respectively, and listed in Appendix 12.2.4.12.

In total, 22 (57.9%) patients took allowed concomitant medications study treatment. The most commonly administered allowed concomitant medications were sodium bicarbonate (4 [10.5%] patients), paracetamol (3 [7.9%] patients), metoclopramide (3 [7.9%] patients) and zopiclone (2 [5.3%] patients). Other allowed concomitant medications were taken by 1 patient each.

Three patients started taking disallowed concomitant medications, which were CYP3A4 inducers/inhibitors, during study treatment for the treatment of AEs (see Table 11.1.15): these were PPD (amlodipine to treat hypertension), PPD (verapamil and diltiazem to treat atrial flutter), and PPD (dexamethasone for asthma and anorexia starting Day 15 of the study [day of last PK sample]). Data for the affected treatment period were excluded for patients PPD since addition of these medications was thought may have affected the PK results.

Modulators of CYP3A4 that were started or stopped prior to treatment or during treatment, could have affected the PK analyses. For the purposes of deciding which patient data to include in the PK analysis set, the following conventions were used. Data for patients receiving a mild or moderate CYP3A4 modulator prior to study start for sufficient time that
Clinical Study Report
Drug Substance AZD0251
Study Code D5160C00009
Edition Number 1
Date 17 August 2015

6.4.4 Relevant medical and surgical history
Listings of relevant medical history and current medical conditions, and relevant surgical history are presented in Appendices 12.2.4.3 and 12.2.4.4, respectively. The medical and surgical history data were as expected for this patient population.

6.5 Use of concomitant medication and treatment compliance
6.5.1 Concomitant medication after study entry
Disallowed and allowed concomitant medications administered during the study are summarised in Tables 11.1.15 and 11.1.16, respectively, and listed in Appendix 12.2.4.12.

In total, 22 (57.9%) patients took allowed concomitant medications study treatment. The most commonly administered allowed concomitant medications were sodium bicarbonate (4 [10.5%] patients), paracetamol (3 [7.9%] patients), metoclopramide (3 [7.9%] patients) and zopiclone (2 [5.3%] patients). Other allowed concomitant medications were taken by 1 patient each.

Three patients started taking disallowed concomitant medications, which were CYP3A4 inducers/inhibitors, during study treatment for the treatment of AEs (see Table 11.1.15); these were NOEATC4L (amlodipine to treat hypertension), 77BB8KD (verapamil and diltiazem to treat atrial flutter), and 36FL21AL (dexamethasone for asthma and anorexia starting Day 15 of the study [day of last PK sample]). Data for the affected treatment period were excluded for patients NOEATC4L and 77BB8KD since addition of these medications was thought may have affected the PK results.

Modulators of CYP3A4 that were started or stopped prior to treatment or during treatment, could have affected the PK analyses. For the purposes of deciding which patient data to include in the PK analysis set, the following conventions were used. Data for patients receiving a mild or moderate CYP3A4 modulator prior to study start for sufficient time that any modulating effects were stable, were included in the PK analysis as long as the treatment remained unchanged through Part A or a change was unlikely to affect the PK results. Data for patients who discontinued their medication in sufficient time for any modulating effects to abate, were also included in the analysis.

- The following 8 patients discontinued CYP3A4 modulators in sufficient time prior to Day 1 or received CYP3A4 modulators well in advance of Day 1 without change in regimen during treatment:
  - FND989V9Z (alprazolam), 9FBPYYLF (fluoxetine hydrochloride), EUW6FXPV (amlodipine), ZPKJ3YDJ (amlodipine), VZBMYFUM (dexamethasone), RP677Y5ER (amlodipine), C3823V5K (ranitidine), ZOCCZ8BYT (ranitidine).
- The following patients started or stopped medication too close to Day 1 or discontinued medication during treatment which could potentially have affected the PK results of the study (also see Section 6.3.2):
8.3.1 Deaths
There were no deaths during Part A of the study.

8.3.2 Serious adverse events
One patient had an SAE. Details are listed in Table 31, and a full narrative can be found in Section 11.4.2.
Clinical Study Report  
Drug Substance AZD9291  
Study Code D3102CO00009  
Edition Number 1  
Date 17 August 2015  

Interstitial lung disease  
Interstitial lung disease was evaluated by review of the ILD and pneumonitis grouped term AEs of special interest. ILD grouped terms. Preferred terms contributing to this grouping were: interstitial lung disease, lung disorder, pneumonitis, diffuse alveolar damage, pulmonary fibrosis, alveolitis, idiopathic pulmonary fibrosis, acute interstitial pneumonitis, and pulmonary toxicity. 

No ILD grouped terms were reported during the study.  

8.3 Deaths, serious adverse events, discontinuation of investigational product due to adverse events, and other significant adverse events  
Summary tables and listings pertaining to this section are presented in Section 11.3 (Tables 11.3.3.1.1 to 11.3.6.2.1 and Appendices 12.2.7.1 to 12.2.7.5).  
Data pertaining to deaths are summarised in Table 11.3.3.1.1 and listed in Table 11.3.3.2.1. Data pertaining to AEs with an outcome equal to death are summarised by SOC and PT in Table 11.3.3.1.2 and key patient information listed in Table 11.3.3.2.2.  
Data pertaining to SAEs are summarised by SOC and PT in Table 11.3.4.1.1 (all SAEs) and Table 11.3.4.1.2 (SAEs causally related to AZD9291). Key information for SAEs is listed by patient in Tables 11.3.4.2.1 (all patients) and Table 11.3.4.2.2 (outcome other than death). A listing of SAEs is presented in Appendix 12.2.7.4.  
Data pertaining to patients who discontinued IP due to an AE (DAE) are summarised by SOC and PT in Table 11.3.5.1.1 (all DAEs) and Table 11.3.5.1.2 (DAEs causally related to AZD9291). Key information for DAEs is listed by patient in Table 11.3.5.2.1.  
Other significant AEs are summarised by category in Table 11.3.6.1.1 and by SOC and PT in Table 11.3.6.1.2. Key information for OAEs are listed in Table 11.3.6.2.1. No OAEs were defined for AZD9291 in Part A of this study.  

8.3.1 Deaths  
There were no deaths during Part A of the study.  

8.3.2 Serious adverse events  
One patient had an SAE. Details are listed in Table 31, and a full narrative can be found in Section 11.4.2.  

In brief, a 65-year-old male patient 7889EAD with a history of hypertrophic cardiomyopathy experienced a CTCAE grade 2 SAE of atrial flutter on Day 1 of the study. ECG had been normal at screening (Day -17) and at Day -3. The patient had ECG findings referred by the investigator as being "tachycardic" (150 beats per minute [bpm]) from 1 hour pre-dose on 04 March 2015 (Day 1) until 12 hours post-dose. These ECGs were reviewed by a cardiologist who confirmed atrial flutter with a 2:1 atrioventricular conduction but presented
Table 11.3.4.2.2 Serious adverse events - Listing of key information for SARS with outcome other than death (Safety analysis set)

<table>
<thead>
<tr>
<th>Subject</th>
<th>Age</th>
<th>Preferred treatment</th>
<th>Treatment period</th>
<th>AEID</th>
<th>[a]</th>
<th>Max time to onset</th>
<th>Max time to treatment</th>
<th>Action</th>
<th>Reasonable possibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>45</td>
<td>ATRIAL 80 mg</td>
<td>AED2931 [a]</td>
<td></td>
<td>1</td>
<td>1/On trt 80 mg</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[a] Calculated for AE(s) starting after the discontinuation of study treatment.
[b] Outcome: RECOVERED = Recovered / Resolved; RECOVERING = Recovering / Resolving; REM = Recovered / Resolved with sequelae; NRNR = Not recovered / Not resolved; P = Fatal.
[c] Actions taken: DN = Dose not administered; DINC = Dose increased; DR = Dose reduced; DINT = Dose interrupted; DPC = Dose permanently discontinued.
[d] As assessed by the Investigator.

Includes adverse events with an onset date between the date of first dose and up to and including 30 days post-last dose of study medication (and prior to enrolment into Part B).

CTCAE = Common Terminology Criteria for Adverse Events (version 4.0).

Table creator: www.d-wise.com
Table 11.3.4.2.2: Serious adverse events - Listing of key information for SAEs with outcome other than death (Safety analysis set)

<table>
<thead>
<tr>
<th>Subject</th>
<th>Term as event of AEs (MedDRA)</th>
<th>Time from start of treatment to last dose</th>
<th>Time from start of treatment to max.</th>
<th>Reasonable possibility</th>
<th>Sex/ age at onset (years)</th>
<th>Investigator term</th>
<th>Time from onset to becoming imm</th>
<th>Received</th>
<th>Action</th>
<th>AE causally related</th>
<th>Treatment sequence - Paused/Fed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miao</td>
<td>ATRIAL FLUTTER</td>
<td>1/On trt 80 mg</td>
<td>1</td>
<td>2</td>
<td>RHCVPR/YES NA</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*(a) Calculated for AEs starting after the discontinuation of study treatment.*

*(b) Outcome: RECOVERED = recovered / resolved, RECEIVING = recovering / resolving, REMAINS = not recovered / not resolved, F = fatal.*

*(c) Actions taken: DMC = dose not changed DMC = dose increased DR = dose reduced DR = dose interrupted DPC = dose permanently discontinued.*

*(d) As assessed by the investigator.*

Age - at study entry

IP = Investigational product NA = Not applicable NR = Not recorded.

Includes adverse events with an onset date between the date of first dose and up to and including 30 days post-last dose of study medication (and prior to enrolment into Part B).

CTCAE = Common Terminology Criteria for Adverse Events (version 4.0).

MedDRA version 17.0.