Date: 13 November 2019

ANONYMIZATION REPORT FOR

LOKELMA® (sodium zirconium cyclosilicate)

Health Canada Control No. 218799

AstraZeneca Canada Inc.

1004 Middlegate Road

Mississauga, Ontario L4Y 1M4

TABLE OF CONTENTS

1. ANONYMIZATION METHODOLOGY

AstraZeneca utilized a mix of qualitative and quantitative risk assessment for this submission. The scope of this submission package as identified by Health Canada under the proactive release process is to include the module 2 documents in scope of the publication initiative and several Clinical Study Reports:

- ZS-002
- ZS-003 EU
- ZS-003 US
- ZS-004
- ZS-004E
- ZS-005
- ZS-006
- ZS-009

AstraZeneca seeks to utilize measured risk and anonymization beyond redaction whenever technically possible. Our approach to anonymization seeks to ensure protection of the individual patient level data, while retaining as much usefulness to readers as possible given the risks of public disclosure. Our process is aligned with the process set forth by Health Canada in the guidance for the Public Release of Clinical Information.

AstraZeneca has utilized an off the shelf product called Blur to measure risk and anonymize the clinical documents for this submission. Blur risk measurement is based on the statistical theory of K-anonymity. For more information on K-anonymity, please see reference in Section 7. AstraZeneca validated the Blur software following industry best practices and standards prior to use for submission delivery.

The rationale for AstraZeneca's anonymization approach is based on several publications and emerging industry best practices. In the effort to balance clinical utility and data privacy, AstraZeneca must also comply with its legal and ethical obligation to protect patient privacy as data stewards. The anonymization methodology described in this report is based on best practices that have been derived from AstraZeneca experience publishing clinical documents, HIPAA laws, PhUSE recommendations, and other industry best practices in the area of Anonymization of individual patient level data sets. For more details about the industry best practices which are the basis of our methodology, please see references in Section 7.

2. IDENTIFICATION OF DATA VARIABLES (DIRECT AND INDIRECT IDENTIFIERS)

In this section we identify and classify the variables as per step 1 of the guidance.

Aggregate data will be retained in most cases throughout the documents, with the exception of when it would lead to the identification of an individual in the way it is presented. For example, minimum, maximum, and median values represent individual values which are often outliers and, therefore, most easily associated with an individual. We are using a qualitative approach for addressing aggregate data at this time.

Individual patient-level (IPD) demographic information was retained when possible.

Due to the sensitive nature of narratives and the likelihood the story can identify an individual, narratives will be removed in full for studies with 100 patients or less. This is defined as the conservative rules as noted in the detailed table below. Please see details at the end of the document on how we will deal with sensitive terms.

In the case of patient deaths, we will redact the details of the death when there are less than 11 deaths in a study. (Applying the conservative rules) For example, within a details table, if it is noted that a particular patient died, we redact “death” as an event or SAE, etc. Also, all details will be redacted from death narratives. All information about the dead subject’s medical history will be redacted, as well as the AEs. Please see details at the end of the document on how we will deal with sensitive terms. Aggregate level information regarding deaths will be retained (with an exception for the case of single death reported in the study. This means if only one subject died on the study and formation related to this death is available on summary level it will be redacted in summary text or/and tables to not disclose what is protected on individual level through facts cross verification among all documents in the submission package revealing this information).

Adverse events and medical history details provide an opportunity to tell a very patient specific story. Thus, we have listed a set of rules used to identify when this information must be redacted to reduce the serious risk of identifying an individual. We are using a qualitative risk assessment approach in this section because we currently don’t have the ability to measure risk on these AEs and sensitive terms.

Any information that is associated directly or indirectly with a person’s identity can be used to build a story that eventual leads to knowing the identity of a single person.

Assignment to treatment like templates of subject identification numbers are redacted to ensure success of patient number scrambling technique throughout the Clinical Study Reports (CSRs).

This is also relevant to site numbers, which are included in patient numbers and therefore redacted when found separate

The following sections provide classification of the direct and indirect identifiers.

2.1 Directly-Identifying Variables

Direct Identifiers are always anonymized in the documents per the Applied Rule	**Applied Rule**
For study participants:	
All Investigator details when associated with a single study participant *See Hrynaszkiewicz reference in Section 7 regarding the risk of a site location leading to identification of an individual study participant*	Redacted
All Individuals: including but not limited to the sponsor, sponsor staff, site staff, investigators, CROs, IRB, Ethics Committees, etc. (when NOT associated with a study participant):	
Personnel Name, Scientific Title, Position and Department	Redacted
Company Name/ Organization Name/Study Site Name/ Hospital Name	Retained
City, State, Country and Zip Code	Retained
Address and other contact details (any references to street number, street name, building names, district, campus, park, phone numbers, emails, fax)	Redacted
Signatures and Handwritten Names	Redacted
All hand-written study related signature dates	Retained
All Individuals when associated with a patient – Staff, Doctors, PIs, Signatories, IRBs, Med Monitors, CROs, CRAs, EC members, etc.	
Personnel Name, Title and Department	Redacted
Company Name/Organization Name/ Study Site Name/ Hospital Name	Redacted
Address and other contact details (any references to street number, street name, building names, district, campus, park, phone numbers, emails, fax)	Redacted
City, State, Country and Zip Code	Redacted
All hand-written text signatures, and/or initials	Redacted
All hand-written study related signature dates	Redacted

* Retained information is noted in green text for ease of differentiation.

2.2 Indirectly-Identifying Variables

However, to manage risk to patient privacy, AstraZeneca have also redacted in the clinical reports several indirect identifiers utilizing an automated redaction tool according to the rules listed in the table below. It is noted that the application of the rules is different, which is based on the approach to measuring and managing risks. This is described in Section 3 of this document:

Indirect Identifiers of Study Participants	**Standard Rules**	**Conservative Rules (100 or less patients)**
Individual Patient metrics (non-summary) * If any of these data elements are not available for risk simulation, then the rule is REDACT		
height, weight, and BMI	Based on the risk simulation	Based on the risk simulation
Sex	Based on the risk simulation	Based on the risk simulation
Age at baseline	Based on the risk simulation	Based on the risk simulation
Race and Ethnicity	Based on the risk simulation	Based on the risk simulation
Country when associated with an individual	Based on the risk simulation	Based on the risk simulation
Site when associated with an individual	Redacted	Redacted
Address Details - This shall include any or all of the following: any reference to street number, street, city, state, and zip code found in reference to a patient shall be redacted. This also applies to partial addresses.	Redacted	Redacted
Aggregate/summary level data		
height, weight, and BMI	• Retain Mean and Standard Deviation (SD), • Redact Min/Max if Risk Sim advised • Median will be retained if the total number of patients in the study population is >= 11, otherwise, it will be redacted	
Sex	Retain both numeric value and classification type when in summary tables and text.	
Age at baseline	• Retain Mean/ Standard Deviation (SD) for Ages • Redact Min/Max • Median will be retained if the total number of patients in the study population is >= 11, otherwise, it will be redacted	

Indirect Identifiers of Study Participants	Standard Rules	Conservative Rules (100 or less patients)
Race and Ethnicity	Retain in tables and text.	
Country	Retained	Retained
Dates		
Month and day related to the individual patient treatment	Offset	Offset
One-time dates such as, hire date, date of degree, date of death, etc of trial participants, including year.	Offset	Offset
Birthdate	Redacted	Redacted
Years (without Month and Day) related to the individual patient treatment	Redacted	Redacted
Dates not related to study participants (for example, study report dates, study milestones including First Subject Enrolled, Last subject completed, First Subject First Visit, First Blood Sample Collected, Last Subject Last Visit)	Retained	Retained
Sites Numbers		
Site Numbers	Redacted	Redacted
Individual Outcomes and specific Adverse Events		
All adverse events, serious adverse event, and associated information that meet any one of the following criteria are redacted to ensure that we are protecting individual privacy: 1- Sensitive, i.e. An associated Stigma to having done that or having that disease 2- Visibly identifiable to someone on the street in a permanent fashion 3- Unique to the population at large 4- Direct reference to relatives and friends is redacted to respect privacy of those individuals Examples of events, but not limited to the following conditions: HIV, Suicide, Pregnancy, Miscarriage, Abortion, Amputations, Deformations, etc.	Sensitive Terms Redacted	Narratives Redacted in Full
Medical Histories	Sensitive Terms Redacted	Redacted in Full
Handling of Deaths		
All details of the death, including associated medical histories, when there are less than 11 deaths in a study.	Sensitive Terms, all medical events and medical history Redacted	Death Listings Fully Redacted but not removed as they are in scope

Indirect Identifiers of Study Participants	Standard Rules	Conservative Rules (100 or less patients)
All details of the death, including associated medical histories (when deaths >=11).	Sensitive Terms Redacted	Death Listings Fully Redacted but not removed as they are in scope
Details presented in aggregate (specifically not associated with an individual and not considered to fall into the category of sensitive Individual outcomes and specific AEs) regarding deaths	Retained with the exception of the case one death on the study (if present then redacted on summary level)	Retained with the exception of the case one death on the study (if present then redacted on summary level)

* Retained information is noted in green text for ease of differentiation.

3. MEASUREMENT OF DISCLOSURE RISK

In this section we explain our approach to risk measurement for this package.

3.1 Reference Population

It is the sponsors responsibility to protect the identity of the study participants. Therefore, AstraZeneca uses the study population as the population for which we are measuring risk for both qualitative and quantitative risk assessment approaches.

3.2 Context of the Data Release

The context of the data release for Public Release of Clinical Documents will result in the posting of these reports in the public domain to the Health Canada's Clinical Information Portal. There are limited restrictions in place around the use of this data. This open release is taken into account in the overall risk management approach.

3.3 Potential Adversary

We assume that a potential adversary would know of an individual within the study as a starting point for any attack. It is widely assumed by sponsors that the release of individual information into the public domain has a very high likelihood of being attacked. Therefore, for public release of the Clinical Study Reports (CSRs) and associated information, we assume the risk that an attacker will try to identify an individual is highly likely.

3.4 Risk Threshold

When measuring risk, we use a risk threshold of a maximum of .09 for the Probability of a Successful Attack (PoSA). This risk threshold is commonly acknowledge publicly by Health Canada and risk experts as a maximum recommended threshold.

Data is transformed based on the risk assessment performed. The risk assessment considers multiple anonymization techniques for key demographic details which are considered industry best practices such as numeric banding opportunities and grouping options. When looking at the demographics we consider retention of gender as most important for scientific utility, thus weighting this variable as more important in risk simulations, compared to the other demographics that are analyzed. The specific demographics analyzed by the risk simulations are included below in the Risk Simulation Details section for the study.

3.4.1 Risk Simulation Details for each study

We utilized the SAS data set of the study participants to analyze the demographic information and measure risk of reidentification for an individual utilizing the anonymization tool Blur.

We selected the following risk simulations per study to use to anonymize the documents

Study	Identifier type	Method	Simulation Hierarchy	PoSA
ZS-002	Age/Age	Numeric Band (CSR - Adult/Child)	CSR -Age	0.026
	Ethnicity/Ethnicity	Drop	CSR - Ethnicity	
	Race/Race	Drop	CSR - Race	
	Sex/Sex	No change	CSR -Gender	
	Country - for participant	No change	CSR - Country	
ZS-003 EU	Age/Age	Numeric Band (CSR Alignment Standard)	CSR -Age	0.014
	Ethnicity/Ethnicity	No change	CSR - Ethnicity	
	Race/Race	Drop	CSR - Race	
	Sex/Sex	No change	CSR -Gender	
	Country - for participant	Drop	CSR - Country	
ZS-003 US	Age	Numeric Band (CSR Alignment Standard)	CSR -Age	0.014
	Ethnicity	No Change	CSR - Ethnicity	
	Race	Drop (Redacted)	CSR - Race	

	Sex	No Change	CSR -Gender	
	Country – for participant	Drop (Redacted)	CSR - Country	
ZS-003 Subgroup Analysis	Age	Numeric Band (CSR Alignment Standard)	CSR -Age	0.014
	Ethnicity	Drop	CSR - Race	
	Race	No change	CSR - Gender	
	Sex	No change	CSR - Ethnicity	
	Country – for participant	Drop	CSR - Country	
ZS-004 and BR-01519-report with Amendments	Age	Numeric Band (CSR Alignment Standard)	CSR -Age	0.033
	Race	Drop	CSR - Race	
	Sex	No change	CSR - Gender	
	Ethnicity	No change	CSR - Ethnicity	
	Weight	Drop	CSR -Weight	
	Country – for participant	Drop	CSR - Country	
ZS-004E	Age	Numeric Band (Age - CSR Alignment)	CSR -Age	0.056
	Race	Drop	CSR - Race	
	Sex	No change	CSR - Gender	
	Ethnicity	No change	CSR - Ethnicity	
	Country - for participant	Drop	CSR - Country	
ZS-005 and br-01670-zs-005-clinical-bioanalytical report	Age/Age	Even Distribution Band (0,20)	Even Distribution Age Banding	0.077
	Location/Country	Drop	Country - Low Frequency	
	Ethnicity/Ethnicity	No Change	Ethnicity - Standard	
	Sex/Sex	No Change	Gender - Standard	
	Race/Race	Drop	Race - Standard	

ZS-006	Age/Age	Drop	CSR -Age	0.030
	Anthropometric-Data/BMI	Drop	CSR -BMI	
	Ethnicity/Ethnicity	Drop	CSR - Ethnicity	
	Sex/Sex	Drop	CSR - Gender	
	Race/Race	Drop	CSR - Race	
	Country - for participant	No Change	CSR - Country	
ZS-009	Age/Age	Numeric Band (CSR Numeric Band - 15)	CSR -Age	0.037
	Race/Race	Drop	CSR - Race	
	Sex/Sex	No change	CSR - Gender	
	Ethnicity/Ethnicity	Drop	CSR - Ethnicity	
	Country - for participant	No change	CSR - Country	
Module 2				No Risk Sim
Patient PD and PK/PD study reports				No Risk Sim
ISS	Age/Age	Numeric Band (CSR Alignment Standard)	CSR -Age	0.010
	Race/Race	Drop	CSR - Race	
	Sex/Sex	No change	CSR - Gender	
	Ethnicity/Ethnicity	No change	CSR - Ethnicity	
QTC	Age/Age	Drop	CSR -Age	0.010
	Race/Race	Drop	CSR - Race	
	Sex/Sex	No change	CSR - Gender	
	Ethnicity/Ethnicity	No change	CSR - Ethnicity	

3.5 Anonymization of Directly-identifying Variables

Additional Data Transformation includes:

- Patient IDs: patient IDs are scrambled to create unique identification numbers which are then used throughout the CSR. This allows the ability to see what happens with one individual, while removing the link between the original study participant numbers.
- Date Offsets: Each patient is allocated a random offset between a minimum and maximum number, which is not disclosed. This number is then added to any dates that the system can accurately associated with that patient. This technique cannot be applied to date ranges, only individual dates at this time. All dates belonging to a patient shall change by the same amount of days. Each patient may receive different offset numbers. This means the integrity between patients' dates remain the same, but the original dates will no longer remain in the document, minimizing the risk that these dates could be matched with other publicly available information such as that posted on social media or available from medical facilities.

4. DATA UTILITY CONSIDERATIONS

In an effort to deliver data utility to the users of these documents, summary level data has been retained wherever possible, as described in Section 2 of this report.

There are cases where specific, sensitive terms or groupings were removed, even from summary level data, to ensure protection of the individual subjects. For protection of the individual subjects AstraZeneca will not specify the specific exact list of sensitive terms removed from the studies within this submission or any other submission.

AstraZeneca understands that optimal clinical utility has been compromised when publishing these redacted Clinical Reports in favour of safeguarding personal information. We encourage researchers who would like to analyze the data associated with this or any AstraZeneca sponsored study to consider making a request for data through our data request portal located at https://astrazenecagroup-dt.pharmacm.com/DT/Home/Index/.

4.1 Precision in support of data utility

Precision is the measure of the amount of damage done to the data we aim to retain or anonymize in the document as defined by the Blur software tool. Higher precision scores indicate less damage to the subject identifiers used in the risk simulation. Specifically, the risk simulation considers the demographics analyzed during risk simulation for creating the equivalence classes, and how much this information would be damaged through anonymization (such as redacting completing, aggregating, implementing age ranges, etc.) based on a selected risk simulation. The

sponsor considered the precision of the data when selecting the anonymization profile described in Section 3.

Precision Details for each study are included in the below table:

Study	Precision
ZS-002	0.381
ZS-003 EU	0.472
ZS-003 US	0.472
ZS-003 Subgroup analysis	0.472
ZS-004 and BR-01519-report with Amendments	0.479
ZS-004E	0.548
ZS-005 and br-01670-zs-005-clinical-bioanalytical report	0.538
ZS-006	0.067
ZS-009	0.528
Module 2	No Risk Sim
Patient PD and PK/PD study reports	No Risk Sim
ISS	0.515
QTC	0.515

5. DEVIATIONS

5.1 Formatting

Due to the use of an automated tool to perform this work, the annotations are found compliant to the format requirements created for EMA Policy 0070. As agreed with Health Canada during the Process Initiation Meeting, these formats were utilized for this submission.

Below is a table of the formats used to mark anonymization throughout the clinical study documents:

HC Definition	HC Preferred format	Package Format
Commercial Business Information	CBI	CCI
Personal Information	PI	PPD
Out of scope pages included within the package	Not Specified	Noted in the document with the text "Out of scope of EMA Policy 0070 Phase 1 – including the out of scope section reference details "

5.2 Addresses

Although not deemed to be Protected Information by the agency, AstraZeneca assumes that street level address details can be used to locate individuals. Addresses of sites are not typically disclosed on registries such as clinicaltrials.gov at the street address level. We do not support disclosure of addresses and do not believe they are necessary for any clinical utility. Therefore, as described in Section 2 and per AstraZeneca's Anonymization Methodology, street level address details which allow the location of an individual's place of work, deemed by the agency not to be Personal Information, is redacted to ensure that individuals cannot be found by location.

6. ATTESTATION

I, PPD on behalf of AstraZeneca Canada Inc., certify that the Anonymization report has been prepared as per the Guidance made available by Health Canada.

PPD

Signed: November 13, 2019

PPD

PPD Regulatory Affairs and Quality Assurance

AstraZeneca Canada Inc.

7. REFERENCES

Health Canada Public Release of Clinical Information: guidance document.
https://www.canada.ca/en/health-canada/services/drug-health-product-review-approval/profile-public-release-clinical-information-guidance.html

HIPAA Regulation: What is Considered Protected Health Information Under HIPAA? HIPAA Journal 2018.
https://www.hipaajournal.com/what-is-considered-protected-health-information-under-hipaa

Hrynaszkiewicz I, Norton ML et al. Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. BMJ 2010:340(7741);304-307.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2825513/pdf/1745-6215-11-9.pdf

Sweeney L. k-Anonymity: A Model for Protecting Patient Privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. 2002;10(5):557-570.
https://epic.org/privacy/reidentification/Sweeney_Article.pdf

Tucker K, Branson J et al. Protecting patient privacy when sharing patient-level data from clinical trials. BMC medical research methodology. 2016;16(supp 1):5-14.
https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-016-0169-4

Winpenny G, Bamford S. De-ID: ‘The What’ and ‘The How’ of de-identifying your clinical trial data. PhUSE Conference publication 2015.
http://www.lexjansen.com/phuse/2015/sd/SD06.pdf