

WHITE PAPER:
**DEVELOPMENT AND IMPLEMENTATION OF A
PHARMA-COLLABORATIVE LARGE HISTORICAL
CONTROL DATABASE**

Authors:

Prakash Bhuyan, Pfizer, Inc.; Chao Chen, GlaxoSmithKline; Jigar Desai, Pfizer, Inc.; Christopher Hart, AstraZeneca; Leah Helvering, Eli Lilly; Daniel Jia, Actavis; Jessica Lim, GlaxoSmithKline; Jiandong Lu, Janssen; Jingyi Liu, Eli Lilly; Stephane Munier, Bristol-Myers Squibb; Josephine Wolfram, Astellas; Edward Bowen, Pfizer, Inc.



TABLE OF CONTENTS

1. INTRODUCTION	5
2. DESCRIPTION OF PLACEBO/STANDARD OF CARE DATABASE	7
2.1. Design	7
2.2. Data Access	8
2.3. User Authorization	9
2.4. Data Sharing/Provision Agreement	9
2.5. Secondary Use of Data	9
2.6. De-identification of Data	10
2.7. Data Elements To Be Included	10
2.8. Inventory Assessment and Loading Priority	12
2.9. Data Mapping to Standards	12
3. APPLICATIONS OF PLACEBO/STANDARD OF CARE DATA SHARING	13
3.1. Use Case #1: Enhanced Safety Signal Interpretation	13
3.1.1. Anticipated Clinical Enhancement	13
3.1.2. Description of Approaches to Utilization	15
3.2. Use Case #2 a & 2b: Control Arm Substitution	17
3.2.1. Anticipated Clinical Enhancement	17
3.2.2. Description of Approaches to Utilization	18
3.3. Use Case #3: Precision Powering	18
3.3.1. Anticipated Clinical Enhancement	18
3.3.2. Description of Approaches to Utilization	19
3.4. Use Case #4 Inclusion/Exclusion Criteria Optimization	19
3.4.1. Anticipated Clinical Enhancement	19
3.4.2. Description of Approaches to Utilization	20
3.5. Use Case #5: Disease Modeling Capabilities	20
3.5.1. Anticipated Clinical Enhancement	20
3.5.2. Description of Approaches to Utilization	21
3.6. Use Case #6: Improved Understanding of Geographic Standard of Care Differences	21
3.6.1. Anticipated Clinical Enhancement	21
3.6.2. Description of Approaches to Utilization	21
3.7. Use Case #7: Biomarker Development	22
3.7.1. Anticipated Clinical Enhancement	22
3.7.2. Description of Approaches to Utilization	23
4. IMPLEMENTATION CONSIDERATIONS	24
4.1. Change Management	24
4.2. Personnel Considerations	24
4.3. Process Redesign	24
4.4. Technology Considerations	24
4.5. Success Measurement (Metrics)	24
5. CONCLUSION	26
6. REFERENCES	27
<i>APPENDIX A. DEFINITIONS</i>	28

LIST OF TABLES

TABLE 1: PROPOSED PSOC DATABASE USE CASE APPLICATIONS	6
TABLE 2: STUDY LEVEL METADATA FOR INCLUSION IN THE PSOC DATABASE	10
TABLE 3: PATIENT LEVEL DATA FOR INCLUSION IN THE PSOC DATABASE	11
TABLE 4: TRANSCCELERATE PSOC DATABASE INITIAL INVENTORY ASSESSMENT	12
TABLE 5: PROBABILITY ANALYSIS FOR OBSERVATION OF AN EVENT IN THE PSOC DATABASE	15

LIST OF FIGURES

FIGURE 1: USE OF HISTORICAL DATA AS DETERMINED BY A MEDLINE SEARCH	5
FIGURE 2: EXAMPLE PROCESS FLOW WITHIN A COMPANY FOR USE OF THE PSOC DATABASE	8

ABSTRACT

Clinical development modernization efforts have become essential as clinical trials have experienced increased expectations, costs and design complexity. The utilization of historical clinical data can enhance drug research and development by refining study design, conduct and analysis. Therefore, TransCelerate BioPharma (TransCelerate) is seeking to develop a database containing placebo and standard of care (PSoC) data from completed clinical trials with the aim to enhance innovative drug product development by better informing clinical safety interpretation and trial design. For purposes of the PSoC database, Placebo data is defined as any data generated from a control arm of a trial whereas the subject received only an inert substance. Standard of care data is defined as any data generated from a control arm of a trial whereas the subject received an active treatment. The PSoC database will be a collaborative data platform composed of clinical data that would be actively contributed by TransCelerate member companies. To enable analysis interpretability, the PSoC database will contain demographic, safety and medical (e.g., comorbidities) data on tens of thousands of subjects from placebo and active control arms of approximately 150 completed clinical studies within the first 3 years of the database building period. The need for drug product development optimization was highlighted by the US FDA launch of the Critical Path Initiative project (FDA 2004). As described by Pocock, the use of historic data can provide a reasonable alternative in cases where the demographic and prognostic factors are understood, where data standards are uniform and where the eligibility criteria and evaluations are similar (Pocock 1976). This white paper provides guidance on the potential applications of this large historical PSoC database and provides examples for the possible implementation of historic data in seven specific applications (use cases).

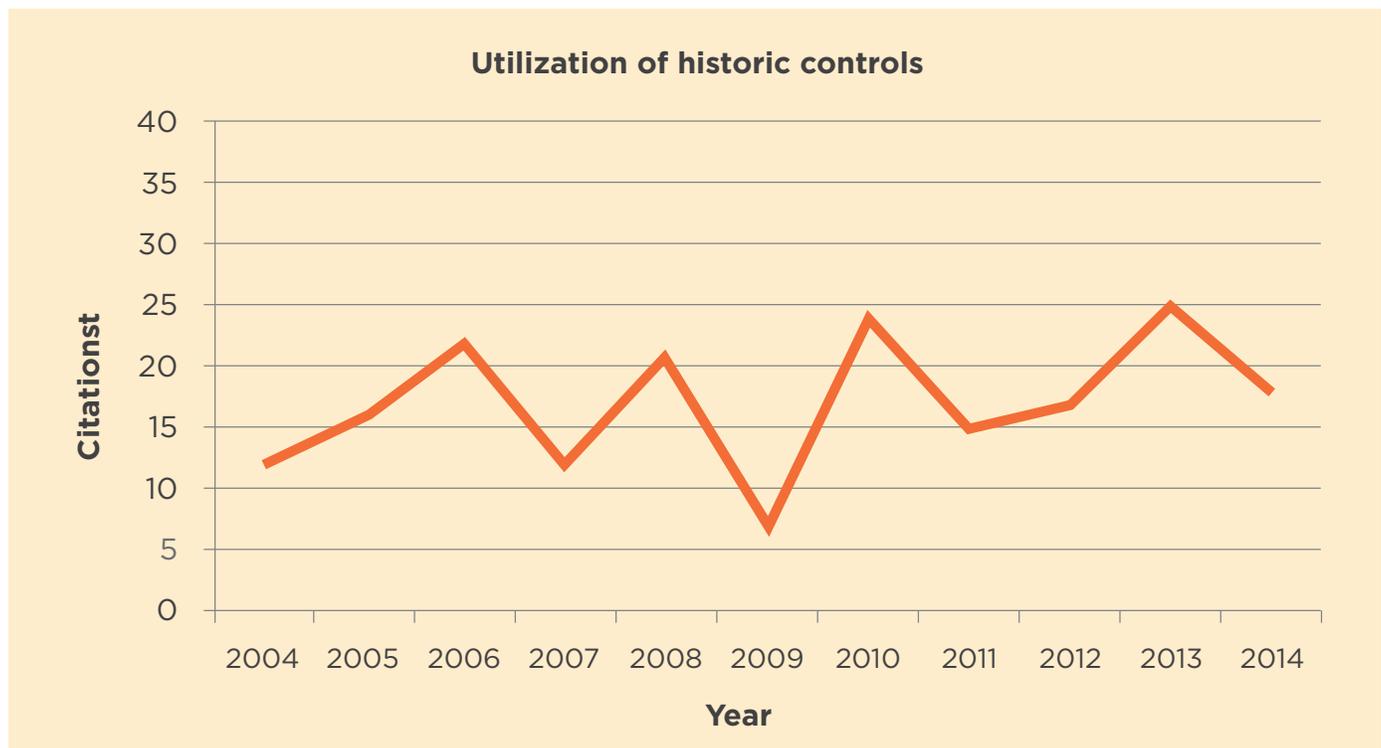
1. INTRODUCTION

The use of historical data in clinical development has been variable and primarily relegated to rare diseases and device development, which reflects its underutilization as a tool in clinical research. A keyword search in the Title and Abstract fields of Medline for the use of historical controls between 2004 and 2014 illustrates the variable use of it across 189 studies to improve study design or data interpretation ([Figure 1](#)). The use of historical control data is expected to improve with enhancements in data structure and standardization and through initiatives such as those being proposed by TransCelerate.

Randomized controlled trials (RCTs) represent the standard for evaluating new drug products

due to the ability to address unanticipated bias that may confound the treatment effect due to imbalances of a variable that impacts the outcome between treatment groups. However, RCTs are resource intensive, time-consuming and may be logistically and analytically limited in cases of rare diseases. With the proper understanding and matching of study design and demographic parameters, historical data can be used in a supplementary manner to help reduce the number of prospective control subjects required during clinical development. Historical data is typically applied to establish data-driven statistical margins for non-inferiority design studies. Therefore, with an understanding of the qualities of the data, historical data can also be appropriately applied to optimize drug product development in general.

Figure 1: Use of Historical Data as Determined by a Medline Search



+Medline search conducted on 6Jan2015 using keyword term 'historic control'

The key aim of the placebo and standard of care (PSoC) database would be to add novel capabilities to enhance clinical development in multiple therapeutic areas, including rare diseases. A number of clinical trial databases have already been established (i.e. Vaccine Adverse Event Reporting System (VAERS), National Institute of Mental Health (NIMH), large insurer-based databases, hospital record-based databases). However, several of these established databases have been limited due to the heterogeneity of the input study data, passive reporting and recording or the relatively small scope of Good Clinical Practice (GCP)-compliant clinical trials included. Initiatives such as the Clinical Data Study Request (<https://www.clinicalstudydatarequest.com/>) allow researchers to request the individual clinical trial data, but do not allow a facile solution to review multiple, large historical studies. Therefore, the PSoC database will incorporate a large volume of GCP-compliant clinical trial data (comprised of both placebo and standard of care) to facilitate the development of innovative drug products. The PSoC database will initially include completed clinical trials since January 2008 to enable better uniformity in data standards and appropriately permissive informed consent use language. The PSoC database is anticipated to continue to grow, as member companies continue to contribute data and will contain demographic, safety and medical (e.g. co-morbidity) data from both placebo and active controls. Initially, approximately 150 completed clinical studies are anticipated to be included within the first 3-year period of database development with anticipated refreshing of data to occur over time as standards of care evolve. The long term goal of the PSoC database is to span a diverse set of therapeutic areas. Criteria for a study's inclusion into the database are described in [Section 2](#).

The composite data from clinical trials of placebo and established drug products represents a large and underutilized body of data from which product development can be enhanced and facilitated. The common concern that has limited the wider use of historic data has been

concern regarding Type 1 error (detecting an effect that is not present). However, approaches can be taken to optimize data interpretability and lower the likelihood of Type 1 error. The PSoC database will be designed to easily understand the demographic and design characteristics of the historical studies. Also, studies included in the PSoC database will be prioritized based upon having established data standards. In cases where more than one historical control study is able to be used, an analysis of intra-study variability can be conducted to understand the variability around the point estimate for the variable of interest. An additional potential data interpretation tool within the PSoC database is the ability to evaluate the variable of interest for placebo-only populations as compared to active treatment populations. Therefore, the utilization of large clinical trial-based data should be pursued and is characterized for the PSoC by seven unique applications (use cases). These use cases are described in [Table 1](#) and are discussed further in [Section 3](#). Additional identified applications would be encouraged and assessed for feasibility.

Table 1: Proposed PSoC Database Use Case Applications

Use Case	
1	Enhanced Safety Signal Interpretation
2a	Control Arm Substitution <i>(Early Phase Trials)</i>
2b	Control Arm Substitution <i>(Late Phase Trials)</i>
3	Precision Powering
4	Inclusion / Exclusion Criteria Optimization
5	Disease Modeling Capabilities
6	Improved Understanding of Geographic Differences
7	Biomarker Development

2. DESCRIPTION OF PLACEBO/STANDARD OF CARE DATABASE

For purposes of the PSoC database, placebo data is defined as any data generated from a control arm of a trial in which the subject received only an inert substance. Standard of care data is defined as any data generated from a control arm of a trial in which the subject received an active treatment. These data may be derived from blinded and open-label trials. Data from investigational products or investigational indications will not be included because the safety and efficacy information for investigational products is considered inconclusive with typically more limited data.

The PSoC database will be implemented in stages.

- » Stage 1 of the database will be a database of control arm data from multiple clinical studies derived from contributing companies. The contributed data would initially include the diabetes, ulcerative colitis, rheumatoid arthritis, chronic obstructive pulmonary disease (COPD), asthma, cardiovascular disease, schizophrenia and vaccines therapeutic areas and would be expected to become available for use by the end of 2015. These particular therapeutic areas were chosen due to significant unmet medical need and/or near-term establishment of data standards in collaboration with Clinical Data Interchange Standards Consortium (CDISC).
- » Stage 2 of the database will provide data compliant with existing CDISC standards to enable better data consistency and interpretability and would be expected to be available by the middle of 2016.
- » Stage 3 of the database will provide support for clinical trial modelling activities and would consider a wider range of therapeutic areas compliant with respective standards and would be anticipated at later date to be determined.

Of note, as a future use, data within areas that do not have established data standards would be allowed in recognition that this data sharing platform may be beneficial for the sharing of small amounts of non integrated data for rare diseases in order to enhance development of treatment for those disorders.

The PSoC database will be developed with full consideration of informed consent requirements. It is expected that each company contributing data will confirm compliance with its organizational policies for the re-utilization of data. Data de-identification will be performed in adherence with privacy laws (HHS 2012). The data within the PSoC database will only include previously published or otherwise disclosed information, representing a meta-analysis approach as often done in medical and scientific fields.

Registered, eligible users from the participating TransCelerate member companies can access the database through secure internet connections, select data to use in their work and download the data for local analysis. Consideration for the potential use of online analytics and the scope of these needs are to be determined as part of Stage 3. All searches within the database will be recorded for audit purposes and assessment of utility. All authorized users must adhere to terms and conditions in their use of these data. Appropriate use access will be governed by appropriate security measures (e.g. login ID, robust password rules and two-factor authentication methods).

2.1. Design

The system will be designed to support researchers from the member companies in using the data for research purposes described by the 7 use cases. To that end, the database will support the search for relevant data and the downloading of selected data for authorized users.

2.2. Data Access

The PSoC database will be established as a collaborative platform for clinical data sharing. A principle of reciprocity will be applied, allowing only contributing members to use the database. This principle should support the ongoing contributions of clinical data for TransCelerate members who participate in the initiative and establish precedent for members who wish to benefit from access to the database once it is active to further encourage the active contribution of data.

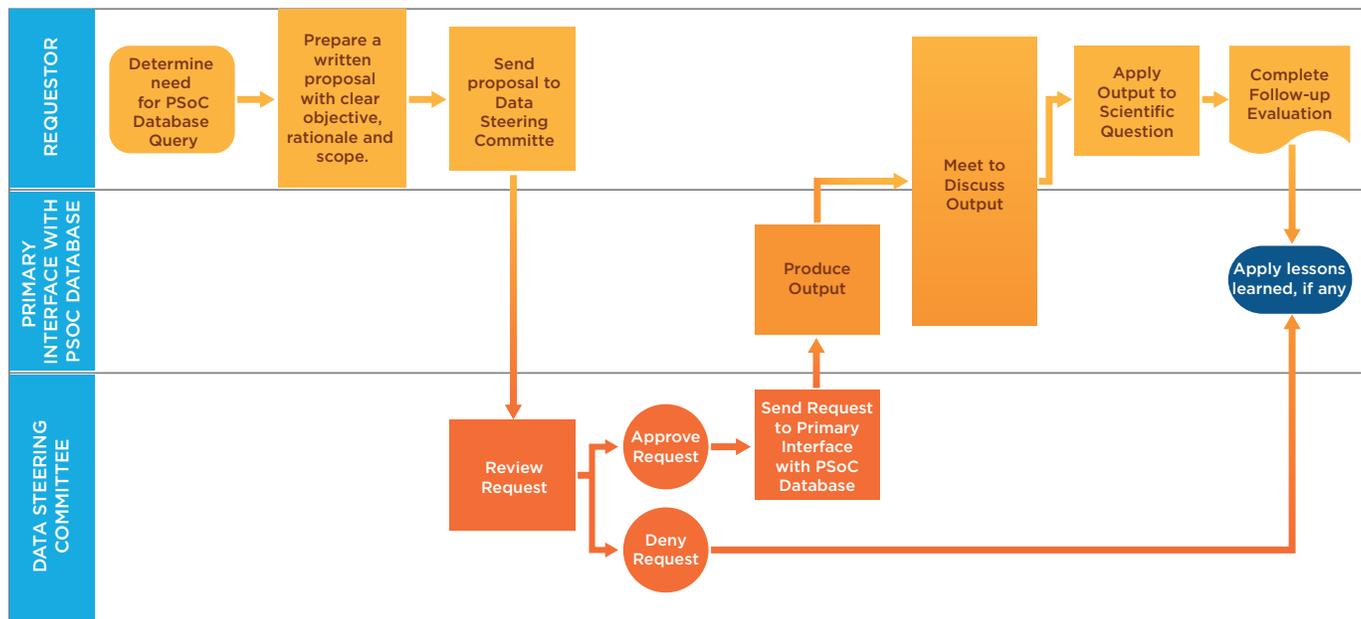
Participating companies will provide the available study data to a formally contracted, independent vendor for storage on the platform. Availability of data requires that participating companies:

- » Execute a clinical data sharing agreement
- » Contribute clinical data to the collection

The database design is currently framed around providing easy access to the participating TransCelerate member companies. All interactions with the data will be logged and at the end of a session, the user may choose to download the data. For the member companies, the user will be prompted to anonymously provide a record of what data is downloaded, when and for what purpose in order to help understand areas of need and potential improvement.

The PSoC database interface would strive to be simple and facile. Each user would be expected to develop a process for work requests, analysis, output delivery, results discussion and continuous improvement based upon the appropriate SOPs and organizational structure. [Figure 2](#) illustrates an example process map that describes how a particular company user may work through their internal organization to make the best use of the PSoC database.

Figure 2: Example Process Flow within a Company for Use of the PSoC Database



Authorized users are allowed to select, download and analyze the data of interest from the PSoC database. Downloading of data is essential for this system in order to support independent research and to enable users to apply user-specific research tools. Data download will be essential for Stage 1 of the system as complex data refining procedures, standardization of data and inter- and intra-domain joins may be needed to make the data suitable for some uses. In Stage 2, the data will be integrated by mapping data to standards (e.g. CDISC) and harmonized to current versions of vocabulary (e.g. MEDRA) to ensure data consistency. Stage 3, pending TransCelerate approval, will require further assessment of user requirements to support the sophisticated activities the PSoC database will need to support, but is anticipated to offer study design modeling tools. Online analytical functions are envisioned to be available as part of Stage 3. These may enhance searching and permit basic visualization to support data selection (e.g. ranges of baseline laboratory results, ages, sex). Access to non-TransCelerate members is out-of-scope for the initial stages of development of the PSoC database, but would be discussed predicated upon the initial applications.

2.3. User Authorization

User authorization will require self-registration and subsequent validation. Only researchers with a valid data-contributing member company or member company specified email address will be able to complete the registration process. Once registered, users must agree to the terms and conditions of use.

2.4. Data Sharing/Provision Agreement

Governance of the PSoC database will be addressed through a formal data sharing agreement that must be executed by all participating members.

The data sharing agreement will describe the supply of clinical data for the collection, the storage of data and the access to data for participating members. Principles addressed would include agreement that:

- » Contributing members would ensure that all data from the contributing company are in accordance with the contributing member's standard operating procedures with regard to compliance with study-specific Informed Consent Documentation
- » De-identification principles applied for the PSoC database will adhere to HIPAA guidelines (HIPAA guideline for de-identification [HHS 2012]). Placebo data and SoC data will retain company and study-level identifiers to facilitate the data substitution application, use case 2 (described in [Section 3.2](#)).

In addition, the data sharing agreement would ensure that users of the PSoC database comply with data sharing principles that include the following:

- » No attempt will be made to re-identify study participants from the data in the collection
- » Data will only be used by the user and others governed by the agreements
- » Data will be used in accordance with the logged research activity

2.5. Secondary Use of Data

A log of use of the database with a statement of purpose would be requested to track use and enable TransCelerate to focus on areas of enhancement and improvement. Secondary (other than for the purposes of the 7 use cases) use of placebo and/or SOC data would be allowed but should be logged as a unique research activity.

2.6. De-identification of Data

To reduce the risks of re-identification of individual study participants from the data in the collection, data will be de-identified to remove appropriate personally identifiable information (PII), in accordance with HIPAA and emerging industry standards (HHS 2012). To help the early provision of data and increase efficiency in the endeavor, companies may supply raw data to the vendor. The vendor will de-identify according to the TransCelerate clinical data transparency model approach in addition to any further de-identification rules of the donating company.

2.7. Data Elements To Be Included

The data will include de-identified subject-level data for study participants in the control arm(s) of each study and select data domains from each study would be included in collection. In Stage 1, it is likely that the differences in data standards between studies will mean that there is diversity of data elements in the collection. Similarly, the de-identification procedures may mean that some data elements appear to be present but may contain unusable values where redaction has taken place to preserve participant privacy or for other reasons.

[Table 2](#) and [Table 3](#) illustrate the study level metadata and patient level data that will serve as the default requirement for any dataset that is submitted for inclusion in the PSoC database.

Table 2: Study Level Metadata for Inclusion in the PSoC Database

Data Element	Use Case(s) Supported
Study Identifier <i>Clintrials.gov number</i>	All – especially if used with regulators
Name of SoC	1, 6
FSFV & LSLV (if not the date, the month)	All (gives indication of how old the data set is in cases of no study identifier)
Therapy Area/Medical Condition	All
Phase of Study	All
Inclusion Criteria	All
Exclusion Criteria	All
Study Objective(s)	All
Name of or class of active treatment	1
Randomization Ratio	All
Primary Endpoint Definition	All
Secondary Endpoint Definitions	All
Visit Schedule	2 – 7
Laboratory Normal Range (inclusive of units) <i>Indicate local or central</i>	All
Redacted Protocol <i>Note: Could be in place of several of the above in a dataset</i>	All

The numbers for the use cases refer to the numbers in [Table 1](#).

Table 3: Patient Level Data for Inclusion in the PSOC Database

Data Element	Use Case(s) Supported
De-identified Visits Dates	
Using relative trial visit days	2 - 7
Date of Birth	All
Gender	All
Ethnicity	All
Race	All
Subject Country	All
Medical History	All
Primary Endpoint Results (including Baseline)	All
Secondary Endpoint Results (including Baseline)	All
Other Baseline Results (e.g. Disease severity)	All
Adverse Events	All
Prior and Concomitant Medications	All
Standard of Care	All
Vital Signs	All
Clinical Laboratory Values	All
Supporting Diagnostic Data (ECG, MRI, tumor assessment, etc.)	All
Physical Exam	All
Treatment Duration	All
Compliance Inclusive of Dosing	All
Trial Status (e.g. Discontinued or Completed)	
Include reason for discontinuation	2 - 7

The numbers for the use cases refer to the numbers in [Table 1](#).

2.8. Inventory Assessment and Loading Priority

The initial selection of studies that will be included in the first stage release will be based on several criteria including:

- » Having better established data standards
- » Participating company nominations of diseases of interest
- » Participating company priority of each disease
- » Applicability of study data to TransCelerate use cases
- » Randomized, controlled studies
- » A last participant last visit date later than January 1, 2008
- » Parallel group design

In October 2014, this assessment produced a list of these diseases of interest and studies for potential inclusion in the PSoC database. The list was further refined to indicate the first 50 trials that would be targeted for inclusion in the PSoC database as shown in [Table 4](#).

Table 4: TransCelerate PSoC Database Initial Inventory Assessment

Disease	Approx. # of Eligible Studies	Approx. # of Subjects in Control Arm
Alzheimer's Disease	12	3,681
Asthma	4	1,795
Cardiovascular Disease	4	10,361
COPD	4	4,249
Diabetes	6	1,015
Rheumatoid arthritis	4	1,702
Schizophrenia	7	1,022
Ulcerative colitis	4	910
Medically Stable Subjects	5	83,778
Total	50	108,513

The list of eligible studies represents an estimate of the largest possible set of data available. Ultimately, the studies submitted will be contingent upon the information sharing policies and standard operating procedures of the participating companies.

Consideration has also been given to the inclusion of large studies with medically stable populations to better support Use Case 1. Use Case 1 focuses on safety analysis and thus depends upon the availability of relevant safety data domain standards. As such, large studies with medically stable subjects across a wider set of therapeutic areas will support Use Case 1 since the disease standards would not be necessary as efficacy data would not be required.

2.9. Data Mapping to Standards

Stage 1:

In the first stage, data will not be mapped to common standards. Users will have access to the unmapped data with the intention of making data available rapidly.

Stage 2:

For the second stage, a vendor will map the data to common CDISC standards. These data will already be gathered together in preparation for data mapping. The volume of mapped data will increase as more standards become available and more study data are added to the collection.

Stage 3:

The third stage adds specific domains to the standardization to support biomarker discovery and trial design modeling capabilities which would aid Uses Cases 5 and 7 particularly. Some data, for example, small sets for rare diseases, may be retained as unmapped data.

3. APPLICATIONS OF PLACEBO/STANDARD OF CARE DATA SHARING

This section provides a general guidance on possible implementations of the PSoC database, recognizing that whether and how to implement will be independently determined by each member company on an institutional case-by-case basis in accordance with respective standard operating procedures and organizational structures. Seven use cases are proposed that would utilize such a database to add value to drug development and these are described in the sections below.

3.1. Use Case #1: Enhanced Safety Signal Interpretation

3.1.1. Anticipated Clinical Enhancement

Clinical development delay or failure in Phase 3 can be caused by product safety-related questions. During clinical safety review, safety events of interest typically arise that require an adequate understanding of the type and frequency of expected and unexpected adverse events, specific for the population being studied. Such safety evaluation during medicinal product development requires an understanding of the background incidence of adverse events (AEs) to distinguish expected from unexpected AEs. Lack of context can lead to incorrect interpretations of AE association or non-association. The process of investigating safety findings can result in months of delay to product development, which impacts development resources and cost.

The expected incidence of AEs is often derived from large observational databases, however, observational databases are limited by passive reporting and, in the case of large insurer databases, potential under- and over-recording (coding) of medical conditions. The PSoC database represents a novel supplement to existing traditional databases, by providing data collected prospectively in an active instead of passive approach more characteristic of randomized clinical trials and will therefore provide adverse event incidence rates for medically stable populations and populations with therapeutic area-specific comorbidities to enable better informed product development decisions as to what is expected vs. unexpected. Historical controls have been successfully applied to provide valuable contextual safety information.

The following are examples of use of historical data in clinical development to support product safety:

- » The safety of intravenous equine F(ab')₂ scorpion antivenom was evaluated in 1534 participants across 5 prospective clinical trials in comparison with a historical control study (Boyer et al 2013). The rates of adverse reaction to the F(ab')₂ product were two orders of magnitude lower than the range (up to 75% for early and 81% for late reactions) historically reported with use of minimally refined whole immunoglobulin products. These data were submitted in support of a biologic license application as an orphan drug and highlights the potential use of historical controls to support orphan drug development.
- » Miyamoto et al 2013 used historical control data of the tolerability profile of high dose melphalan as a conditioning regimen for stem cell transplantation treatment of multiple myeloma to provide a contextual safety profile with which to evaluate the safety and tolerability of bortezomib given together with high dose melphalan. They were able to generate pilot data that supported the safe co-administration of these chemotherapeutic agents. This example shows the benefit and application of the use of historical control data for conditions that are relatively uncommon.

The following are examples of use of historical data in clinical development to support product safety and efficacy:

- » Weschler et al 2014 evaluated the safety and efficacy of conversion from 1-2 antiepileptic medications to lacosamide monotherapy using a historical data-controlled, multicenter study design. The authors concluded that lacosamide monotherapy was effective and had a favorable safety profile compared with historical standard of care therapy. This example illustrates the use of historical control data for uncommon conditions.
- » Melichar et al 2013 studied the safety and efficacy of an investigational regimen of bevacizumab and low-dose interferon- α 2a as compared with historical control data for bevacizumab and standard dose interferon. This phase 2 study showed improved tolerability without reduction in efficacy against renal carcinoma.
- » The ePlacebo database was developed as an integrated database comprised of historical data for placebo-only recipient subjects from Pfizer's completed clinical trials completed after the year 2000 and contains studies of medically stable subjects and studies of subjects with unstable medical conditions (Desai et al 2013). The ePlacebo database has been used within Pfizer to provide an estimation of the incidence rate of adverse events in a population that only received placebo. The database contains demographic, safety and medical (e.g. co-morbidities) data on approximately 56,000 subjects from the placebo arms of approximately 1134 studies. Criteria for a study's inclusion into the database were the presence of placebo controlled arms, parallel group design, completion after the year 2000, study duration of at least six weeks and compliance with the Pfizer Data Standard (PDS). Subjects in the ePlacebo database of studies were allowed to receive non-study medications. The ePlacebo database showed that it is technically feasible to pool portions of control arm populations through a stratification and segmentation approach for a comparable large clinical group database.
- » The eControls database was developed as an integrated database comprised of historical data for medically stable subjects from Pfizer's completed clinical trials completed since the year 1994 (Bhuyan et al 2013). The eControls database has been used within Pfizer to provide an estimation of the incidence rate of adverse events in a medically stable population. This database contains demographic, safety and medical (e.g. co morbidities) data on approximately 52,000 subjects from both the placebo and standard-of-care arms of approximately 40 studies from Pfizer's Clinical Trials Information Repository. Studies of subjects with chronic, stable co-morbidities such as hyperlipidemia were included as representative of a population that would be potentially eligible for a prophylactic vaccine trial or a small molecule trial of medically stable subjects. Criteria for a study's inclusion into the database were parallel group design, completion since the year 1994, a study period of at least 6 weeks and compliance with Pfizer Data Standards. Subjects in the eControls database of studies were allowed to receive non-study medications. The eControls database only included studies of small molecule compounds and, therefore, contains data on systemic adverse events, but not local reactions. The eControls database showed that placebo and active control data could be re-purposed to provide valuable adverse event incidence rate safety context.

The PSoC database will be based upon safety data recorded under controlled, clinical study settings and provide much-needed safety context to enable better informed product development decisions.

The expected initial size of the PSoC database would be >10000 subjects included in the first 3 years of database development. Not all datasets would be appropriate for pooling and would need to be guided by the therapeutic area and population of interest. The probability to observe at least 1 event in the PSoC database would be based upon the database size and the true probability of occurrence of the event as shown in [Table 5](#) below.

Table 5: Probability Analysis for Observation of an Event in the PSoC Database

PSoC Size (N)	True Probability of Event Occurrence	Probability to Observe at Least 1 Event in PSoC
100	0.1%	10%
100	1%	63%
100	2.3%	90%
100	10%	>99%
1000	0.1%	63%
1000	0.23%	90%
1000	1%	>99%
10000	0.023%	90%
100000	0.0023%	90%
1000000	0.00023%	90%

Individual therapeutic area datasets are anticipated to range from 100 to 100000 subjects. Datasets of medically stable subjects would have broad applicability across therapeutic areas by providing general safety context for a medically stable population and would be expected to grow to >100000 subjects. The approximate therapeutic area-specific sample sizes anticipated in the first 3 years of the database build are noted in [Section 2](#).

3.1.2. Description of Approaches to Utilization

The application of the PSoC database to enhance safety signal interpretation is expected to have potential impact and use for the Biostatistics, Clinical Development, Regulatory and Pharmacovigilance groups. A safety query would be initiated typically by one of the product development functional areas. Initially, the adverse event term must be clearly defined. Once the general event term is defined, all of the associated terms should be identified using current MedDRA coding. Multiple types of entries for a single category can exist and can cause an underestimation of true incidence if not identified and included. Given that studies of licensed active compounds are part of the PSoC database, it is essential to know if certain classes of medications may influence the adverse event by either a treatment effect (e.g., non-steroidal analgesics reducing the frequency and severity of headache) or by increasing the likelihood of the event of interest (e.g. non-steroidal analgesics increasing the likelihood of gastritis). Once these particular requirements and influencing variables are identified, the PSoC database can be queried with or without certain studies to help interpret the impact of confounding by the active, standard-of care, compound. Furthermore, analyzing subsets of data by important demographic or baseline characteristics could be employed if appropriate for the setting or even more sophisticated subject-matching or adjustment algorithms utilized. It is important to consider that in surveillance studies of a non-clinical trial, general population, the bias of concomitant medications would still be present and not likely to be as easily identified. Once the PSoC database has been queried, the incidence rate of the event of interest can be used as an estimation of what frequency of occurrence might be expected using either placebo and/or licensed active compounds as a benchmark.

As with any database, the PSoC database has certain strengths and weaknesses which influence the interpretability of the results. An understanding of these strengths and weaknesses enables the appropriate utilization of the database and are detailed below.

The notable strengths include:

- » The quality of data used in the PSoC database is required to be in accordance with GCP standards. This represents a significant improvement over traditional databases based upon passive reporting of adverse events and retrospective analyses of medical records or large insurer data, which is often limited by under- and over-reporting issues.
- » As only licensed, standard of care drug products and placebo will be included in PSoC, the database reflects what might be expected and considered acceptable for a placebo or licensed product (depending upon the subsets chosen). Information on active licensed compounds is important because it provides a background rate of what might be expected for compounds that were safe enough to achieve licensure with the consideration that the benefit-risk profiles may differ. Information on placebo is also important as it provides a background rate of what adverse events might minimally be observed in a given population.

The notable challenges include:

- » MedDRA search terms need to be appropriately and comprehensively included, which requires in depth medical knowledge of the general condition.
- » The inclusion of standard of care compounds in the PSoC database introduces two potential confounders, compound-related AEs and therapeutic effects. To help compensate for these confounders, multiple product classes should be included in the data analysis, to prevent a single product class being predominant in the results. As the PSoC database includes more studies of different compounds, treatment-dependent bias will diminish.
- » It is not possible to differentiate between studies that solicited or did not solicit for certain AEs. Studies that actively solicit for events, such as active recording of systemic events (e.g. headache), would be expected to increase the chance of recording such events. Furthermore, the time after initiation of dosing during which certain events may have been solicited can vary between studies and, therefore, may affect reporting.
- » Studies with more efficient recording tools, such as electronic diaries, may have a higher overall AE reporting rate. Electronic recording tools have only recently become widely used and so databases with a higher proportion of studies done in recent years may have higher general reporting of solicited events.
- » Clinical study populations may be more proactive with healthcare and may be inherently healthier than the general population, which would potentially underestimate the rate of AEs. Alternatively, clinical study populations may also be more likely to report events than the general public, which would increase the incidence rates.

3.2. Use Case #2a (Early Phase Trials) & 2b (Late Phase Trials): Control Arm Substitution

3.2.1. Anticipated Clinical Enhancement

RCTs represent the standard for evaluating new drug products due to the ability to address unanticipated bias and avoid type 1 error. However, RCTs are resource intensive, time-consuming and may be logistically and analytically limited in cases of rare diseases. Therefore, it is proposed that, with the proper understanding and matching of study design and demographic parameters, historical data can be used in a supplementary and substitutive manner to reduce the number of prospective control subjects required during clinical development (Pocock 1976). A few advantages of utilizing the historic data are listed below:

1. Acceleration of development

Substitution of control arm subjects with appropriately matched historical control data can reduce the total number of control subjects required, shorten the duration of study and thereby speed up clinical development enabling effective medicines to reach patients more quickly.

2. Enhancement of understanding of investigational product

Unequal randomization may be used to place proportionally more subjects in the experimental treatment arm in a study potentially increasing the relative amount of information both on the efficacy and safety of the novel treatment.

3. Minimizing unnecessary exposure to the PSoC arm (Clinical Equipoise)

Borrowing historical control arm data allows fewer subjects to be randomized to the SOC or placebo arm, which might be a less effective treatment option. In one example, historical data were used as a comparator to evaluate whether the rate of maternal to infant vertical transmission of hepatitis B infections was comparable for more recently developed thimerosal-free hepatitis B vaccines and thimerosal-containing hepatitis B vaccines, which are no longer available for use. This study showed that the effectiveness of the thimerosal-free vaccines was similar to that of the older formulations and illustrates an example of using historical controls when a prior standard of care is no longer available (Doherty et al 2009).

4. Optimization of study power

For RCTs, it is critical to control type 1 error and maintain reasonable power for a given hypothesis of interest. If historical control data is borrowed appropriately, such that it is nearly identical to the prospective control, then a more precise estimate of parameters of interest, such as tumor response rate, percentage of Hemoglobin A1C reduction from baseline, etc. can be obtained by incorporating the historical data. As a result, the study power could be improved without changing other parameters. Type 1 error can be minimized (Viele et al 2014).

5. Enabling innovation for treatments of rare diseases

For some rare disease areas with small subject populations, it remains challenging to enroll subjects to conduct a RCT. Historical data has been effectively used to develop innovative therapies in rare disease areas. In one example, Weschler, et al evaluated the safety and efficacy of conversion from 1 or 2 anti-epileptic medications to lacosamide monotherapy using a historic-controlled, multicenter, study design (Weschler et al 2014). The authors concluded that lacosamide monotherapy was effective and had a favorable safety profile compared with historic standard of care therapy. In a second example, historical placebo data were used to assess whether lithium carbonate slowed the progression of amyotrophic lateral sclerosis in a Phase 2 trial and showed that the rate of progression was not impacted. For orphan indications involving rare diseases, access to subjects for control groups is limited and there are ethical concerns regarding randomizing subjects to suboptimal placebo arms. Baseline features for the historical control group and active treatment group were matched and included an established ALS functional rating scale (Miller et al 2011).

6. Optimized use of resource for clinical development

With fewer patients in a trial, overall resources and costs to execute each trial would be lower. At the study level, fewer patients result in fewer clinical visits, fewer expensive imaging/lab tests and fewer investigator sites required to conduct research. At a broader research and development level, this allows resources to become available for underfunded projects to enable broader development of medicines.

Overall, utilizing the historical data can result in more accurate point estimates, increased power and reduced type I error in clinical trials, which facilitates drug development.

3.2.2. Description of Approaches to Utilization

The application of the PSoC database to enhance study design and the interpretation of the study outcomes is expected to have potential impact and use for Biostatistics, Clinical Development, Regulatory and researchers. In a typical workflow stream, a research team would assess the need for incorporation of historical controls into the proposed study. Consideration should be given to whether there is sufficient historical information available in the PSoC database and/or in the literature for the relevant control from relatively recent studies of comparable design, conduct and subject population. The PSoC database would then be queried to determine if the appropriately matched historical control data exist. For studies intended for support of registration, discussions with regulatory agencies about the acceptability of this approach should occur well in advance, as early as the Pre-phase 1 stage and no later than at the End of Phase II meeting with a review of the proposed application of the PSoC historical data within the product development plan and a review of the characteristics of the proposed historical control population.

An *a priori* systematic literature review should be performed to be able to appropriately identify the relevant, representative clinical trials within the PSoC database to be included as historical controls. (Higgins and Green eds 2011). Comparability of the historical controls to the concurrent control (i.e. with respect to time period differences between the studies, changes in standard of care, studies run in different countries, different entry criteria, etc.) should be assessed and methods should be implemented to weight/discount the historical information (Viele et al 2014, Hobbs et al 2013 and Dane and Wetherington 2014). Additionally, Gamalo et al discussed order restricted power priors and how they can be used to incorporate different levels of evidence, e.g., double-blind studies could be assigned a larger weight than open-label studies (Gamalo et al 2014). Pocock described

conditions that a historical trial must meet before it can be used in a substitutive manner (Pocock 1976). An approach to limiting the potential for bias with the incorporation of historical controls would be to provide estimates based upon a range of weights/discounting in order to assess the impact of the historical data on the current study results.

3.3. Use Case #3: Precision Powering

3.3.1. Anticipated Clinical Enhancement

Missing data in RCTs can have critical impact on the quality and validity of findings from interventional studies. Dropouts and missed visits by study participants can impact intent-to-treat analyses and threaten the ability to draw causal inferences. Moreover, depending on the therapeutic area and study design, regulatory agencies can classify subjects who dropout before specified time points as failures, irrespective of the reason of withdrawal (Hughes et al 2012). Despite efforts to mitigate the effects of missing data, the great majority of RCTs will have missing data. In recent years, there has been considerable effort in developing statistical methods to handle missing data. However, when data are missing and downstream statistical methods are applied, there are assumptions, usually unverifiable, that make conclusions from such analysis sensitive to the assumptions and analytical method utilized (Carpenter et al 2007). It should be noted that although great strides have been made in analytical techniques, these methods are simply the best available option rather than solutions (Dziura et al 2013). The PSoC database would enable analyses, with more fidelity, to uncover predictive factors that improve retention rate of subjects, allow for larger numbers of subjects completing study visits and mitigate chances of sensitivity analyses findings that lead to conflicting results. Operationalizing these findings would bring increased efficiency and improved clarity and quality to clinical trials and increased confidence in study results from regulators. The data within the PSoC database would enable the identification of factors that could reduce missing data in clinical trials.

3.3.2. Description of Approaches to Utilization

Currently, there is a focus on predicting the expected proportion of missing data rather than identifying predictive factors to reduce missing data. Given the importance and the gains in quality and efficiency by attenuating missing data, a more proactive approach during the design phase to minimizing missing data of clinical trials is needed (Dziura et al 2013 and Hughes et al 2012). Several reports in the literature, including those from regulatory agencies (FDA), have called for a paradigm shift from downstream handling of missing data to strategies that reduce missing data (O'Neill and Temple 2012). A recent study meta-analysed both the SOC and experimental arms from pooled historical HIV clinical trial data to investigate which subject subgroups were more likely to drop out for potentially avoidable reasons unrelated to treatment. The study found that in a typical HIV Phase 3 non-inferiority trial, 5% fewer dropouts could have increased the trial's power from 90% to 95% thus reducing risk of type II error or reduced the sample size by 15% while maintaining power at 90% (Hughes et al 2012), which allows development resource allocation to be optimized. The problem of missing data in clinical trials is universal and given that no single method/analysis will be definitive in managing missing data, the most impactful approach is to take steps during the study design phase to prevent missing data.

3.4. Use Case #4 Inclusion/Exclusion Criteria Optimization

3.4.1. Anticipated Clinical Enhancement

The definition of the study population is fundamental to data interpretability and data generalizability and directly impacts the enrollment parameters of a study. If the study population is defined appropriately, it has a greater chance of revealing the true causal relationship between the intervention therapy and the consequence (i.e., effect). However, understanding the optimal study population for a new intervention is often challenging to clinical trial designers.

The historical control data of the PSoC database enables a means to choose study inclusion and exclusion criteria in a better informed manner improving study feasibility, probability of success and data interpretability. The relevant demographics, comorbidity, baseline disease characteristics, concomitant medications and main efficacy and safety outcomes data from the placebo and/or standard of care control arms will enable individual drug developers to explore and identify the potential correlations between subjects' characteristics and clinical efficacy or safety outcomes. Consequently, trial designers would gain prior knowledge of study subjects with respect to which subjects have higher placebo responses, which subjects are likely to have harmful disease deterioration without treatment and what comorbidity conditions or background medications are more likely to be associated with serious adverse effects and could dilute the treatment effect. This learning exercise opportunity will enable trial designers to learn from the past study samples of the intended study populations and choose more feasible, appropriate and reliable study population eligibility criteria for a new clinical trial. Through analyses of clinical outcomes of previous trials, drug researchers and developers could uncover a substantially greater effect in a specific subset of the overall study population and provide a basis for studying that subset in a new clinical study, either as the only targeted study population or as the primary endpoint population in a study of a broader population (FDA 2012). The large scale, shared PSoC database will also enable clinical trial designers to provide better justification in lieu of the historical data for the specified inclusion and exclusion criteria of a new clinical trial protocol, providing prior evidence to IRBs, regulators and study investigators on the defined study population for the new clinical trial. Clearly, appropriately choosing the inclusion and exclusion criteria for a new study will impact directly upon trial recruitment (enrollment), whose feasibility can be guided by historical recruitment statistics. Furthermore, utilization of the large scale, shared PSoC database would enable trial designers to determine the inclusion and exclusion criteria more appropriately for the designed trial treatment duration and to avoid or minimize the number of protocol amendments which generally increase product development costs.

3.4.2. Description of Approaches to Utilization

The application of the PSoC database for this use case is expected to involve clinical development, biostatistics and potentially clinical operations. Depending on the diseases of interest, users will be able to search the PSoC database for patient level data on the efficacy and safety outcomes and factors such as:

- » Demographic characteristics of the prospective subjects expected to benefit from the new intervention therapy
- » Baseline disease status or disease stage requirement of the prospective subjects who have the potential to benefit from the new intervention
- » Baseline efficacy measures and concomitant medication use requirement
- » Shifts in the disease population over time, which make it uncertain as to whether the selected characteristics of the study population reflect the current disease population
- » Risks to participants in longer-term clinical trials (e.g., cancer trials or heart disease trials over more than 1 year)

Analyses based on the extracted information can identify the factors that may influence the efficacy and/or safety outcomes of interest. Collections of inclusion and exclusion criteria across trials can be collated by the users to investigate the differences for certain groups of criteria (e.g., lab exclusion). Additionally, the past enrollment pattern (e.g., enrollment curve over the duration of the trial) can be analyzed by varying the inclusion/exclusion criteria based on disease characteristics.

Unique to this use case is the knowledge management for the lists of inclusion and exclusion criteria. Because of the nature of free text information, mapping similar criteria across trials is challenging, if not impossible. A keyword search capability of the system will be important once the PSoC database procures more studies for each of the diseases of interest.

Guided by the results of these quantitative analyses and qualitative investigations, clinical teams can fine tune the inclusion/exclusion criteria to optimize study interpretation, enrollment challenges and the balance between enrichment and generalizability of the study populations.

3.5. Use Case #5: Disease Modeling Capabilities

3.5.1. Anticipated Clinical Enhancement

The ability to use the PSoC database to model trial design would enable better informed decision making for trial design options and minimize associated costs (as opposed to simply conducting a 'traditional' trial). Overall trial design quality would be improved, which would enable a higher probability of success and reduce the need for protocol amendments. This would be anticipated to impact Phase 2 through Phase 4 trial design.

Drug-disease model based design and analysis would have significant impact on Phase 2a (proof-of-concept) and 2b (dose ranging) trials. Disease-model-based design and analysis would reduce equivocal or failed Phase 2 trials and provide more confidence in design, subject stratification and dose selection for Phase 3. At least 1 amendment is implemented in 69% of all protocols and 39% of these amendments are considered 'somewhat' or 'completely' avoidable; such amendments add cost and time to development program (estimated at \$450K and 2 months of implementation time) (Getz et al 2011).

The use of the PSoC database for clinical trial modelling is limited to the portfolio focus, but therapeutic area models could potentially be developed for the majority of major and rare diseases.

3.5.2. Description of Approaches to Utilization

Predictive modeling represents an underdeveloped area that the PSoC database aims to empower. As an example, historical data were used in the development of crizotinib, a cancer therapy. The ability to conduct simulations using relevant biomarkers enabled the prediction of the efficacy outcome.

http://www.accessdata.fda.gov/drugsatfda_docs/nda/2011/202570Orig1s000StatR.pdf

3.6. Use Case #6: Improved Understanding of Geographic Differences

3.6.1. Anticipated Clinical Enhancement

Most studies are now conducted globally. The sites and countries are typically selected based on prior feasibility experience. However, many additional factors should be taken into consideration to optimize recruitment, which can be enabled by use of the PSoC database. For example, clinical trial sites should be placed where the population of interest is most prevalent, which can help reduce sample sizes and recruitment times. In geographies where compliance rates are low or study dropout rates are high, preventive measures could be implemented (e.g. more selective screening, better focused study investigator meetings, frequent phone calls and reminders). Use of the PSoC database to better understand these factors should help increase the speed of recruitment in clinical trials, which in turn means a higher likelihood of on-time regulatory submissions and market access.

By understanding regional differences, the clinical trial designer can also determine if the population being studied is generalizable to the global

population. By better understanding geographic variability, more informative and focused clinical studies within geographic regions would be possible. The PSoC database could also allow better understanding of differences in the control arm effectiveness, safety and overall benefit-risk, by geography. It could also allow optimization of benefit-risk assessment by selecting countries or regions with more predictable placebo and standard of care effect and more accurate safety reporting. This could help increase the probability of success of clinical trials.

Additional important study variables that can vary by geography include:

- » Disease stage
- » Medical diagnostics
- » Medical standards of care (e.g. prior therapies, concomitant medications)
- » Medical practice recommendations
- » Safety reporting processes and standards
- » Compliance to treatment protocols
- » Public perceptions of diseases and treatments

The PSoC database will enable a better understanding of these factors, which should be considered in the design of global product development programs.

3.6.2. Description of Approaches to Utilization

When designing a clinical trial, the population of interest needs to be well defined. Once the population of interest is defined, the PSoC database can be queried for the parameters of interest. [Section 3.4](#) provides a process description to optimize the definition of the population using the PSoC database.

The PSoC database can then be investigated to assess whether the population of interest resides in specific locations. Findings could potentially be cross checked with literature or based on experience from previous studies. This will help to prevent opening sites in geographic areas where target population may not exist.

For the indication and the population of interest, use of the PSoC database will also allow assessment of whether the effect is geography dependent. Efficacy and safety subsets analyses by sites, countries and regions could be performed to understand heterogeneity. Cochran's Q test could also be performed along with calculation of the I^2 value to assess whether study results are consistent across geography. Statistical modeling can also be considered for assessing, for example, whether treatment effect would differ in one region compared to others. It would also allow identifying potential prognostic factors, such as baseline covariates, concomitant medications or background therapy, which may be geography dependent and may impact the effect of interest or the safety reporting. However, these subset analyses results should be interpreted with caution. Ideally, endpoints of interest, regions of investigation and the potential prognostic factors should be predefined to test specific null hypotheses. Otherwise, too much exploration may lead to incorrect conclusions due to multiplicity of statistical testing.

Based on all those elements, the clinical trial designer can make an educated decision on the population on which the clinical trial should be conducted, as well as on the regions/countries where this population will be recruited.

3.7. Use Case #7: Biomarker Development

3.7.1. Anticipated Clinical Enhancement

A biomarker may be any measurement/assessment made on or any observation/expression shown from a biological system. A main aspect of biomarker research relates closely to identifying the trigger of a disease and tracking disease progression or pace (e.g., Breast Cancer, CHF); another important aspect of biomarker research and discovery relates directly to prediction of and/or association with sensitive clinical outcome endpoints (e.g., tumour size reduction versus overall survival improvement). In an ideal world, if a clinical trial were conducted in selected patients with identical characteristics (for instance, disease initialization and progression attributes, biological characteristics, demographics, background

medical conditions, pattern of experiencing adverse reactions, compliance to treatment interventions and adherence to study drug intake, outcome report and study visit assessment schedules), the real correlations between some bio-marker endpoints, patient reported outcome endpoints and endpoints that accurately reflect the effectiveness of treatment intervention would have been quantified/estimated much more accurately. The above ideal setting, however, is infeasible in current clinical trials, particularly because most clinical trials are currently conducted globally, leading to inherited heteroscedasticity between regions or countries in terms of subject characteristics, measures of disease advance/progression, concomitant medication usage, tolerance thresholds of feeling an adverse reaction or standards of compliance to study intervention. Consequently, analytic work to identify (discover) and quantify biomarker endpoints for clinical effectiveness within a single clinical study or even a couple of large clinical studies combined often is strenuous and exhausting with no conclusive findings. With a large scale clinical database like the PSoC database, the drug product developer will have the opportunity to explore the aforementioned differences and more accurately quantify the relationships between biomarkers and clinical outcome endpoints with higher precision and less bias. This will, in turn, enable optimizing selection of the overall study population in a new clinical study across different regions/countries; and this would also help with enriching a region-specific study population for region-specific regulatory approval requirements.

Therefore, development of a clinical biomarker or a set of clinical biomarkers that track disease and predict clinical outcomes is important to the translation of innovation to the public domain and to the revolution of pivotal clinical trials for new medicine. The PSoC database is designed to facilitate the mission of discovery and quantification of clinical biomarkers in drug research and development. Utilization of an established (with full development, use and validation) biomarker in late-stage clinical trials would efficiently enhance the targeted disease population and increase the probability of trial success (FDA 2012)

The PSoC database is expected to initially have disease specific data in the therapeutic areas noted in [Section 2](#) and the laboratory information from these clinical trials that has been already collected, assayed and reported would also be included. Gene expression data may also be included when available. These data represent an underutilized resource for biomarker development. This application of the large PSoC database will improve the understanding of laboratory data, enable the identification of optimal study participants, better understand what is within normal expectations for laboratory evaluations for a given population and ultimately yield more interpretable study result data. When linking together the corresponding baseline laboratory test results, the baseline disease status/stage and baseline clinical efficacy measures, for example, there is a potential to distinguish between different subgroups of subjects based one or more threshold levels of one or more laboratory endpoints that potentially have different responses to an intervention. Over time, the PSoC database is expected to expand the collection of data to other therapeutic areas.

One limitation is that the PSoC database is retrospective and does not enable the prospective analysis of biospecimen data. The second limitation is that subjects in the PSoC database come from different drug developers' clinical trials that may differ in key selection criteria of the enrolled (or randomized) subjects; so data in the PSoC database may not accurately reflect the desired disease population to be studied. Another limitation is that certain therapeutic areas collect information in broader detail, whereas others may not. The initial datasets loaded into PSoC database may have focus upon therapeutic areas that had better defined and established data standards. However, these limitations would not necessarily prevent utilizing the PSoC data for potential marker discovery.

3.7.2. Description of Approaches to Utilization

With the advances of computer-derived informatics and software, many researchers in medical research and drug discovery started to utilize more data for mining biomarkers, understanding disease progression or verifying one or more measures that affect or predict patients' disease course. For example, Atassi et al recently described the Pooled Resource Open-Access ALS Clinical Trials database (PRO-ACT) in the October 8 Neurology online (Atassi et al 2014). They reported that the PRO-ACT database contains over 8600 subjects who participated in 16 Phase 2 or 3 trials and one observational study; and nearly 5000 of those recorded, list the ALS-Functional Rating Scale (ALS-FRS) scores, more than 300 included mortality information while the trial was ongoing and more than 7000 records indicated whether the patient took riluzole (the only treatment for ALS). They reported that using the PRO-ACT database, Atassi and colleagues confirmed suspicions, raised in much smaller trials, that uric acid, creatinine and body mass index affected the ALS disease course in the larger sample; and the higher levels of uric acid and creatinine in the blood predict a slower course of ALS (Atassi et al 2014). In one publication with utilization of the PRO-ACT database, scientists sorted the cases into fast and slow progressors that could help exclude subjects whose disease course is so slow that they would not be expected to benefit from short-term treatment in a typical confirmatory clinical study (Gomeni et al 2014).

4. IMPLEMENTATION CONSIDERATIONS

Implementation will be decided by member companies unilaterally, voluntarily and in accordance with the respective member company standard operating procedures.

4.1. Change Management

The PSoC database presents a tremendous opportunity for the pharmaceutical industry to optimize the manner in which future clinical trials are designed and analyzed. As with any major alteration in the approach to clinical development, there should be focus on impact to people, process and technology. This section provides further recommendations for the PSoC database access and utilization process, in line with the overall process map proposed in [Section 2](#).

4.2. Personnel Considerations

Sponsors that adopt this solution must ensure that the appropriate resources are in place to execute the designed process for PSoC data access and utilization. Initially, it is critical that senior leadership is aware and provides endorsement for the new approach. Communication from the senior leadership of the organization of a participating member company throughout the organization will also support the embedding of the change within the organization. It should be fully recognized that the execution of the new process will certainly include a cross-functional team composed of clinical operations, safety and biostatisticians among others. Each Sponsor who elects to implement should consider performing a skillset inventory to assess any gaps in expertise and determine how those gaps, if any, will be addressed to ensure a successful implementation.

4.3. Process Redesign

As previously discussed in section 3, within each participating member organization, a cross functional team composed of representatives from each of the impacted functional areas should convene to develop a customized process map for the organization. Components of the process to be developed include, but are not limited to, determination of explicit roles and responsibilities, governance of the process through creation of new or modification of existing Standard Operating Procedures, use cases that the organization will adopt, lead time necessary to fulfill requests and how results will be interpreted.

4.4. Technology Considerations

In Stages 1 and 2 of the TransCelerate solution, sponsors will retrieve desired datasets from the repository and perform analysis utilizing in-house applications. Therefore, it is critical that an organization gives consideration to those analytic tools that are currently available to them. An inventory approach should be utilized to assess gaps so that they can be filled in a timely matter. In tandem with the process redesign process, analytic tools should also be called out in order to illustrate when and where in the process utilization of such tools will take place.

4.5. Success Measurement (Metrics)

To assess impact and success of the new methodology, it is recommended that metrics be utilized. Metrics would not contain company identifiers and would be explicitly defined such that collection, analysis and reporting are straightforward. In this manner, there will be a clear indication of progress. For each given proposed metric, a historical control going back at least one year should be determined, if applicable, to provide a baseline against which future activity can be measured.

It is fully recognized that there could be general metrics utilized to assess all use cases, however, there will also be others that could be measured to assess the impact of a single use case. Sponsors retain the discretion to determine whether to use metrics and which to use, but if they elect to use metrics should consider

focusing on metrics that align to three major categories; Quality, Cycle Time and Efficiency.

Metrics should be designed and agreed upon during the process redesign efforts and can be modified over time with experience should new ones be identified or existing ones deemed non-valuable.

Utilization and impact metrics will also enable the TransCelerate PSoC to focus on areas of potential future growth and improvement.

Examples of metrics include the following:

Quality

Quantify whether dropout rate, discontinuations, protocol violations are decreasing in respective therapy area and indication as a result of applying this methodology

Cycle Time

Subject recruitment rate

Efficiency

Subject recruitment costs

If member companies include or refer to the PSoC database within publications, the TransCelerate PSoC database would need to be referenced.

5. CONCLUSION

The PSoC database will enable pharmaceutical R&D companies to share clinical data in a noncompetitive, collaborative environment to enhance the development of new medicines. Seven initial use cases for historic clinical trial data are feasible and will help guide and streamline development programs. These data are distinct from other large databases by being GCP-compliant, by containing placebo data subsets and by having the ability to more fully understand the demographic characteristics, which in turn enhances data interpretability. By repurposing clinical trial data, the large resources and costs that have already been invested to conduct past studies can enable better informed decision-making, capture efficiencies in clinical trial processes, reduce costs and empower development of high quality medicines for patients.

6. REFERENCES

1. Atassi N, Berry J, Shui A, Zach N, Sherman A, Sinani E, et al. The PRO-ACT database: design, initial analyses, and predictive features. *Neurology* 2014;83(19):1719-1725.
2. Bhuyan PK, Desai J, Carlsson M, St. Louis MV, Bowen E, Cantor MN. eControls: a model to characterize the incidence of common adverse events in a vaccine-eligible population. Abstract 152. *Infectious Diseases Society of America's IDWeek 2013; San Francisco, CA; October 2-6, 2013.*
3. Boyer L, Degan J, Ruha A-M, Mallie J, Mangin E, Alagon A. Safety of intravenous equine F(ab')₂: insights following clinical trials involving 1534 recipients of scorpion antivenom. *Toxicon* 2013;76: 386-393, 2013.
4. Carpenter J, Pocock S, Lamm CJ. Coping with missing data in clinical trials: a model-based approach applied to asthma trials. *Statist Med* 2002;21:1043-1066.
5. Dane A and Wetherington JD. Statistical considerations associated with a comprehensive regulatory framework to address the unmet need for new antibacterial therapies. *Pharmaceut Statist* 2014;13:222-228.
6. Desai JR, Bowen EA, Danielson MM, Allam RR, Cantor MN. Creation and implementation of a historical controls database from randomized clinical trials. *J Am Med Inform Assoc* 2013;0:1-7.
7. Doherty R, Garland S, Wright M, Bulotsky M, Liss C, Lakkis H, et al. Effectiveness of a bivalent Haemophilus influenzae Type b - Hepatitis B vaccine in preventing Hepatitis B virus infection among children born to Hepatitis B e antigen-positive carrier mothers. *Pediatr Infect Dis J* 2009;28:777-781.
8. Dziura JD, Post LA, Zhao Q, Fu Z, Peduzzi P. Strategies for dealing with missing data in clinical trials: from design to analysis. *Yale J of Bio & Med* 2013;86:343-358.
9. FDA. Guidance for Industry - Enrichment Strategies for clinical trials to support approval of human drugs and biological products. Draft Guidance. December 2012.
10. FDA. Innovation or Stagnation: Challenge and Opportunity on the Critical Path to New Medical Products. March 2004.
11. Gamalo MA, Tiwari RC, LaVange LM. Bayesian approach to the design and analysis of non inferiority trials for anti-infective products. *Pharmaceut Statist* 2014;13:25-40.
12. Getz KA, Zuckerman R, Cropp AB, Hindle AL, Krauss R, Kaitin KI. Measuring the incidence, causes, and repercussions of protocol amendments. *Drug Information Journal* 2011;45(3):265-275.
13. Gomeni R and Fava M. Pooled resource open-access ALS clinical trial consortium: amyotrophic lateral sclerosis disease progression model. *Amyotroph Lateral Scler Frontotemporal Degener* 2014;15(1-2):119-129.
14. Health and Human Services (HHS) Dept of the US Government. Guidance regarding methods for de identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) privacy rule. November 2012.
15. Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.*
16. Hobbs BP, Carlin BP, Sargent DJ. Adaptive adjustment of the randomization ratio using historical control data. *Clinical Trials* 2013; 0:1-11.
17. Hughes S, Harris J, Flack N, Cuffe RL. The statistician's role in the prevention of missing data. *Pharmaceut Statist* 2012;11:410-416.
18. Melichar B, Bracarda S, Matveev V, Alekseev B, Ivanov S, Zyryanov A, et al. A multinational phase II trial of bevacizumab with low-dose interferon- α 2a as first-line treatment of metastatic renal cell carcinoma: BEVLiN. *Annals of Oncology* 2013;24:2396-2402.
19. Miller RG, Moore DH, Forshew DA et al. Phase II screening trial of lithium carbonate in amyotrophic lateral sclerosis. *Neurology*. 2011; 77: 973-979
20. Miyamoto T, Yoshimoto G, Kamimura T, Muta T, Takashima S, Ito Y, et al. Combination of high-dose melphalan and bortezomib as conditioning regimen for autologous peripheral blood stem cell transplantation in multiple myeloma. *Int J Hematology* 2013;98:337-345.
21. O'Neill RT and Temple R. The prevention and treatment of missing data in clinical trials: an FDA perspective on the importance of dealing with it. *Nature* 2012; 91(3):550-554.
22. Pocock SJ. The combination of randomized and historical controls in clinical trials. *J Chronic Dis* 1976;29:175-188.
23. Viele K, Berry S, Neuenschwander B, Amzal B, Chen F, Enas N, et al. Use of historical control data for assessing treatment effects in clinical trials. *Pharmaceut Statist* 2014;13:41-54.
24. Weschler RT, Li G, French J, O'Brien TJ, D'Cruz O, Williams P, et al. Conversion to lacosamide monotherapy in the treatment of focal epilepsy: Results from a historical-controlled, multicenter, double-blind study. *Epilepsia* 2014;55(7):1088-1098.

APPENDIX A. DEFINITIONS

Background Therapy

Medications continued from prior to study enrollment

Coding

Assigning a number to a disease process, surgical procedure or other type of health care service for the purpose of reimbursement, health care planning and research

Compliance

Pre-specified measures of adherence to study protocol activities

Data Anonymization

The removal of information that identifies an individual or can be used to identify an individual and which does not preserve the ability to reconstruct the redacted information

Data De-identification

The removal of information that identifies an individual or can be used to identify an individual and which preserves the ability to reconstruct the redacted information if necessary

Data Elements

The data fields that will be required as part of a member company's submission of a trial dataset to the PSoC database

Data Mapping

The conversion of all data fields within a trial dataset to CDISC standards

Data Segmentation

The practice of dividing a dataset into smaller groups

Data Standardization

Generally accepted, uniform criteria for entry of information

Data Steering Committee

An individual member company's governance body

Data Stratification

The practice of organizing data according to descriptive variables

Harmonization

The alignment of all data fields to same version of dictionary terms especially as it pertains to Medical History, Adverse Events and Concomitant Medication

Heteroscedasticity

The circumstance in which the variability of a variable is unequal across the range of values of a second variable that predicts it.

Historical Data

The legacy trial data that each member company submits to the PSoC Database with Last Patient Last Visit having occurred no earlier than January 2008

Investigational Indication

A new indication that is under development for a licensed product

Medically Stable Subjects

Subjects who may have chronic medical conditions that do not require altering medication regimen or intensive medical attention (e.g. stable medical regimen and controlled condition) and have minimal comorbidities

Member Company

The TransCelerate member that is participating in the PSoC database initiative

Placebo

Refers to any data generated from a control arm of a trial in which the subject received only an inert substance

Primary Interface for PSoC Database

An individual member company's primary owner of the PSoC database

Redacted Protocol

Document that defines study aims and procedures which has had information removed

Requestor

The individual who initiates use of the PSoC database within a member company

Secondary use of data

The use of data for any purpose other than the 7 stated use cases

Standard of Care

Refers to any data generated from a control arm of a trial in which the subject received a marketed, active treatment. This includes study groups which received a placebo and standard of care treatment to maintain study blinding