5 August 2013

To: Department of Health and Human Services, Food and Drug Administration
Re: Docket # FDA- 2013- N-0271

The International Society for CNS Clinical Trials and Methodology (ISCTM) welcomes this opportunity to respond to the FDA request for comment: *Availability of Masked and De-identified Non-Summary Safety and Efficacy Data.*

The ISCTM formed a Working Group to review and provide comments. The International Society for CNS Clinical Trials and Methodology (ISCTM) offers these comments for consideration based on our experience and expertise in human CNS research. The ISCTM is an independent organization focused on advancing the development of improved treatments for CNS disorders. No member of this Working Group received compensation for comments provided. Comments represent personal opinions and not that of the institution, agency, or company affiliation of Working Group members.

Authors (in alphabetical order):

Alvaro Camacho, MD, MPH, *University of California, San Diego*
Phillip Chappell, MD, MBA, *Pfizer, Inc.*
Nicholas DeMartinis, MD, *Pfizer, Inc.*
Robert M. Hamer, PhD, *The University of North Carolina, Chapel Hill*
Richard Keefe, PhD, *Duke University*
Atul R. Mahableshwarkar, MD, *Takeda Development Center Americas, Inc.*
Stephen R. Marder, MD, *Semel Institute for Neuroscience at UCLA*
Cherri Miner, MD, MBA, *Eli Lilly and Company*
Danielle Popp, PhD, *MedAvante, Inc.*
Holly Posner, MD, *Pfizer, Inc.*
Jonathan Rabinowitz, PhD, *Bar-Ilan University*
Jill Rasmussen MD, *psi-napse*
Joanne Severe, MS, *Research Consultant*
Michelle Stewart, PhD, *Pfizer, Inc.*

FDA Questions	**Comments**
(1) What factors should be considered in masking study data (e.g., data fields from regulatory submissions to remove, or modify, number of different products to pool within a product class)?	With the availability of study information on CT.gov, it may be somewhat unrealistic to mask product identity from researchers who are familiar with the indication and product class. Ideally, a study protocol would be provided for each dataset. However, since that is unfeasible, at the least a minimum set of study design facts and figures should be provided, perhaps via a standard protocol description template. A minimal set of information might include: number of subjects, number of treatment groups, inclusion of placebo arm, duration of washout if used, duration of treatment, post-baseline assessment points during treatment, post-treatment assessment points, source of ratings (e.g., self-report, blind on-site assessor, non-blind assessor, etc.), information about missing values (e.g., is data original or are missing values imputed), are raw score item-level data provided or only summary scores, identification of rating instruments and their versions (necessary for interpretation of summary scores), identification of primary and secondary outcomes, and planned hypotheses. The question of what needs to be masked would be a function of the particular question being addressed.
(2) What limitations, if any, should there be on the Agency's ability to make available the masked data as described previously?	Would the availability of such data be proactive (e.g., posted on a public access website) or through requests only? A substantial risk is the use of unsound or inappropriate research approaches by researchers who may be biased, inexperienced, or lacking the necessary resources leading to the potential for distorted or incorrect results that could negatively impact public health. Pooling data across multiple clinical trials, multiple programs and multiple drug development programs to conduct meta-analyses can be vulnerable to artificial results due to multiple factors (e.g. differences in MOA of the compounds, differences in clinical trial design, and differences in patient populations). There should be clear agreed upon guidelines around retrospective meta-analyses of pooled clinical trial data. Scientists expressing interest in accessing masked and de-identified data from clinical trials need to be thoroughly vetted for their history of delivering scientific quality research and their qualifications and resources for conducting the analyses they plan, as well as a detailed research/statistical analysis plan that ensures that its objectives can be met with the requested data and have a low likelihood of distorted findings that could negatively impact public health. In addition to the vetting mechanism that FDA would establish to ensure appropriate use of the data, consideration should be given to establishing

	a public website where analytic plans are posted prior to making data available. This would allow for broader vetting by interested parties, create a user community, and help to hold those who request data to the plans and uses they proposed. To minimize risk for harm from making these kinds of data available, consideration should be given to providing education to potential researchers on aspects of the analysis of these kinds of combined data that can lead to distorted or incorrect findings. Which organization/institution would assume the responsibility for providing guidance and oversight on research conducted with these data? Will there be a data use agreement or contract that will govern the use of data to be shared?
(3) Are there any additional factors FDA should consider in de-identifying data in addition to FDA's requirement to remove any names and other information (e.g., birth date, death date, local geographic information, contact information) which would identify patients or research subjects before disclosing information?	Would the trial data falling under this proposal come from trials that have already been submitted to the FDA or from future trials? The informed consent of the clinical trials conducted to date likely does not include research subject agreement to the disclosure of individual information in a public forum. In addition, the laws of different countries need to be taken into account in releasing individual data from global trials. The data need to be "limited access datasets" that comply with the rules of the HIPAA regulations. Such datasets must be stripped of any variable that might violate the privacy of any participant. Obvious examples of variables that must be omitted are: names of participants or family members; addresses; telephone numbers; social security numbers or their equivalents; place of birth; city of birth; contact data; birth date; name of clinical site; clinician and interviewer identifiers. All occurrences of dates in the dataset must be converted to "days from baseline." Birth date may be converted to age. New, randomly generated identification numbers must replace original identification numbers and the dataset(s) must be re-sorted according to new identification number. Codes linking the new and original data are not to be included in the limited access dataset, but should be kept securely by the investigator/data center in case discrepancies in the data arise that require re-checking. Other sensitive data, or data that may lead to the loss of participant privacy because of the uniqueness or rarity of the data, must also be omitted. For rare diseases and analyses of groups with very small sample sizes where it may be possible to nearly identify an individual subject, particular caution is warranted. Analyses of this type should be restricted as with Medicare and Medicaid data. Sensitive data, including illicit drug

	use, risky behaviors (e.g., carrying a gun or exhibiting violent behavior), sexual behaviors, and selected medical conditions (e.g., alcoholism, HIV/AIDS) must be omitted if the size and focus of the trial are such that knowledge of these variables could lead to loss of participant privacy. Unedited, verbatim responses, stored as text data (e.g., specified in "other" category), must be omitted. Genetic markers and identification of family relationships and pedigrees sufficient for individual identification must be omitted. Variables with low frequencies for some values, that might be used to identify participants, may be recoded, if appropriate. There may be other variables identified by the study centers that may make it easy to identify individuals. Careful consideration should be given to all such variables to have them recoded or omitted.
(4) Would regulatory changes facilitate implementation of such a proposal, and if so, what changes would be most useful?	The majority of requests for this type of data from the EMA have been from commercial entities rather than academic ones, suggesting regulatory changes may be needed to reduce the risk of inappropriate use of masked and de-identified data for commercial or competitive benefit rather than the general welfare intended by this initiative. Clear guidelines and policies on how this would be done, incorporating input from all key stakeholders, would be helpful. Consideration may need to be given to changing requirements for consenting to participate in a study. When patients sign a consent form agreeing to be study subjects, they place tremendous trust in the researchers to use the data for public good in the manner described in the consent. Generally this pertains to the use of their data for that particular study and otherwise held in confidence, with no expectation that their data (de-identified, masked or otherwise) will be made publicly available. If data sharing is to become standard practice, the consent forms for all clinical trials should make this explicit so that subjects are fully aware of how their data will be used. Regulatory changes may also be needed to address handling adverse drug reactions that may be identified during such analyses.
(5) Which situations do you believe disclosing masked data would be most useful to advance public health?	Release of masked data for placebo arms from clinical trials in a given disease area could contribute to optimization of clinical trial design and signal detection without placing intellectual property or commercial interests at risk. Disclosing substantial amounts of masked data for a given therapeutic class with multiple available compounds is more likely to deliver value than releasing data for novel mechanism or first in class compounds. Making masked data available to researchers may best be prioritized

	based on the impact on public health and/or safety but requires input from a variety of stakeholders. Developing a priority list of key topics would be prudent and might be accomplished by establishing a process to get public input on topics which can be reviewed by a group representing public health interests such as researchers, patients, consumers, sponsors, and other interested parties, as was done by AHRQ a few years ago.
Other	This process will require very careful qualification of the potential users, and very careful documentation of hypotheses, plans, and analyses before the users obtain the data. Further, there should be very intensive oversight by the FDA, or a group/body consisting of appropriately qualified people selected by the FDA. These safeguards will help prevent clogging the literature with noise and prevent confusion around real public health concerns. Different large-scale data-sharing models such as the FDA is considering currently exist or are being proposed (e.g., C-Path initiative, EMA practice, the NIH). The FDA should inventory these various models for sharing data and undertake a careful study of the strengths and weaknesses of each, perhaps even engaging with the stakeholders to better understand how the different models operate. This should be done before the FDA moves forward with finalizing or establishing an approach. Due to the wide scope of this proposal and the complicated nature of sharing data on this scale, we strongly urge the FDA to conduct a second round of public opinion gathering prior to finalizing any data-sharing proposal.