



August 5, 2013

Dockets Management Branch (HFA-305)
Food and Drug Administration
5630 Fishers Lane, Rm. 1061
Rockville, MD 20852

Re: Docket No. FDA-2013-N-0271: Availability of Masked and De-identified Non-Summary Safety and Efficacy Data; Request for Comments

Dear Sir/Madam:

The Biotechnology Industry Organization (BIO) thanks the Food and Drug Administration (FDA) for the opportunity to submit comments on the "Availability of Masked and De-identified Non-Summary Safety and Efficacy Data."

BIO represents more than 1,100 biotechnology companies, academic institutions, state biotechnology centers and related organizations across the United States and in more than 30 other nations. BIO members are involved in the research and development of innovative healthcare, agricultural, industrial and environmental biotechnology products, thereby expanding the boundaries of science to benefit humanity by providing better healthcare, enhanced agriculture, and a cleaner and safer environment.

GENERAL COMMENTS:

BIO and its member companies appreciate FDA's commitment to modernizing regulatory science and are committed to enhancing public health through advancements that fill knowledge gaps in drug development. To that end, BIO supports voluntary, responsible sharing of appropriately masked and de-identified safety and efficacy data as one potential way to develop new knowledge or new scientific insights. BIO member companies routinely publish clinical research results and collaborate with academic researchers, while both adhering to applicable privacy laws and regulations and protecting confidential commercial information.

It is important to recognize that issues of transparency and disclosure of regulatory data are extremely complex and currently being debated and evaluated globally in various venues. We urge FDA to consider these issues in the context of other transparency initiatives, including the Agency's own Transparency Initiative, as well as the Expansion of the ClinicalTrials.gov database as directed by the Food and Drug Administration Amendments Act (FDAAA) enacted in 2007. As stated in BIO's April 2009 testimony on the ClinicalTrials.gov expansion, it is critical to ensure that data to be interpreted outside of FDA's expert review process have scientific merit and can enhance the treatment and



safety of patients.¹ Accordingly, the purpose of disclosure of such data, as well as appropriate recipients of such data, must be clearly and transparently identified.

In addition, it is important to strike a balance in achieving dual public health goals, both recognized in the Agency's Notice: the goal of enhancing scientific knowledge to advance public health and patient care, as well as the goal of protecting confidential and trade secret information, in order to maintain the incentives necessary for further investment in and development of new treatments for patients. With regard to the latter, the trade secret and commercially confidential nature of clinical data may remain even when such data are masked or pooled with data from other products.

As FDA contemplates mechanisms to responsibly release data from approved marketing applications, BIO recommends that FDA first conduct a formalized review of voluntary data sharing initiatives, such as the Cardiac Safety Research Consortium², to share lessons learned, to evaluate if these initiatives have been productive and efficient in addressing scientific questions, and to identify the gaps FDA's own initiative could address. BIO strongly recommends that the Agency then consider a pilot program (*e.g.*, data sharing in indications where there is focused research, such as an indication where there have been multiple trial failures, so that research may be conducted to develop a more reliable study design) as a next step.

As FDA develops its plan, it is essential for the Agency to publish for comment a complete, detailed, and transparent proposal for data sharing prior to its implementation. FDA's plan, at minimum, should:

- Clarify the purpose of its data-sharing proposal (*e.g.*, is the purpose to advance regulatory science/fill knowledge gaps through qualified third party research, to assist the FDA in pursuing its own research, or both?). It is imperative that data sharing arrangements not be used as a means to evaluate data to influence regulatory decisions under the FDA's purview or as a way to shortcut regulatory exclusivity and intellectual property protections.
- Clarify the scope of its data-sharing proposal (*e.g.*, applicable only to confirmatory trials for marketed drug, biologic, or device products, regardless of whether results are considered "positive" or "negative"^{3,4}).

¹ BIO Comments on Public Meeting on Expansion of the Clinical Trial Registry and Results Databank (2009) http://www.bio.org/sites/default/files/BIO_Comments_NIH_2009_0002.pdf

² Cardiac Safety Research Consortium (2006-present) <https://www.cardiac-safety.org/>

³ BIO Position Statement on Clinical Trial Registries and Dissemination of Clinical Trial Results (2005) <http://www.bio.org/articles/bio-position-statement-clinical-trial-registries-and-dissemination-clinical-trial-results>

⁴ BIO Comments to NIH on Expansion of the Clinical Trial Registry and Results Data Bank (2009) http://www.bio.org/sites/default/files/BIO_Comments_NIH_2009_0002.pdf



- Clarify effective date for potential availability of applicable data sets (*e.g.*, data sharing arrangements are applicable to data sets held by FDA upon effective date of any future Guidance issued).
- Define the criteria that constitute qualified research, including development of a list of important research categories that are within scope and specification of criteria used to select a product class or indication.
- Define “de-identified data” and parameters to ensure prevention of re-identification and fully protect the privacy of patients. The provision of de-identification must be consistent with informed consent and applicable privacy laws and regulations.
- Define “masked data” and safeguards that will ensure the maintenance of masking efforts, thereby protecting commercial confidential information. Sponsors of approved applications should have an opportunity to address with FDA which particular data would be protected by trade secret/CCI provisions even if masked.
- Ensure reciprocal transparency for Sponsors whose data are being disclosed by:
 - Defining “non FDA expert” qualified researchers who could be granted access to data (*e.g.*, academic researchers, government scientists, etc).
 - Defining the process for evaluating qualified research requests, including processes for evaluating whether the available data can support the research objective, specifying the duration of time that data will be accessible (*i.e.*, limited duration), and requiring research requestors to submit a research proposal for review by FDA and the Sponsor company. Requestors should be required to enter into agreements with Sponsors outlining the scope of research, as well as other parameters and safeguards, prior to being granted access to data sets.
 - Defining a notification process that will ensure Sponsors are notified before their data are released.
- Identify appropriate financial and staff resources necessary to mask, de-identify, and redact data for potential disclosure. The aggregation, standardization and electronic conversion of clinical data for products submitted in paper or legacy formats alone represents a considerable added workload burden on the Agency. In this era of budgetary austerity, it is important that FDA prioritize its limited resources towards its core mission activities.



SPECIFIC 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)?

In the Federal Register Notice, FDA commits to not releasing trade secrets and “business-related confidential commercial information,” such as “information concerning licensing agreements and information identifying suppliers.”⁵ This statement does not define the full scope of proprietary information that is protected from disclosure by federal law and FDA regulations. Further, the notice implies that non-summary safety and effectiveness data, which are otherwise not available for disclosure, lose their status as confidential commercial information and as trade secret when masked and are therefore disclosable. Concluding that masked data are automatically not protected from disclosure may conflict with federal statutes⁶, FDA regulations⁷, judicial precedent⁸,

⁵ 78 Fed. Reg. at 33423.

⁶ Several federal statutes protect non-summary safety and effectiveness data in marketing applications from disclosure by FDA. Specifically, although the Freedom of Information Act (FOIA) generally requires FDA to release requested agency records, exemption 4 of FOIA authorizes agencies to withhold trade secret and confidential commercial information. 5 U.S.C. § 552(b)(4). Further, the Federal Trade Secrets Act (FTSA) prohibits federal employees from disclosing, “in any manner or to any extent not authorized by law,” information that “concerns or relates to trade secrets, processes, operations, style of work, or apparatus.” 18 U.S.C. § 1905. The FTSA broadly protects “virtually every category of business information likely to be in the files of an agency,” including any information that falls within Exemption 4. *CNA Fin. Corp. v. Donovan*, 830 F.2d 1132, 1140 (D.C. Cir. 1987); see also *id.* at 1151 (“the scope of the [FTSA] is at least co-extensive with that of Exemption 4 of FOIA”).

⁷ FDA’s implementing regulations likewise protect confidential commercial information and trade secrets from disclosure. FDA defines confidential commercial information as “valuable data or information which is used in one’s business and is of a type customarily held in strict confidence.” 21 C.F.R. § 20.61(b) & (c). A trade secret is a “commercially valuable plan, formula, process, or device” that is used in biopharmaceutical product manufacture, has a direct link to the productive process, and is “the end product of either innovation or substantial effort.” 21 C.F.R. 20.61 (a).

⁸ Federal courts have interpreted FOIA and FDA’s regulations implementing FOIA to protect from disclosure non-summary safety and effectiveness data as confidential commercial information. The governing *National Parks* test provides that agency records are “confidential” if their disclosure is likely “to cause substantial harm to the competitive position of the [submitter].” *Nat’l Parks & Conservation Ass’n v. Morton*, 498 F.2d 765, 770 (D.C. Cir. 1974). To show this standard is met, the submitter need not show actual competitive injury, but must establish “actual competition and the likelihood of substantial competitive injury.” *Pub. Citizen Health Res. Group v. FDA*, 704 F.2d 1280, 1290 (D.C. Cir. 1983) Applying this test, courts generally have found that “raw research data used to support a pharmaceutical drug’s safety and effectiveness” are confidential commercial information. Department of Justice, Department of Justice Guide to the Freedom of Information Act 326 (2009), available at http://www.justice.gov/oip/foia_guide09/exemption4.pdf These data are also “trade secrets” under state trade secret law and therefore “property” protected by the Fifth Amendment of the U.S. Constitution. See *Ruckelshaus v. Monsanto Co.*, 467 U.S. 986, 1003-04 (1984).



and the Agency's longstanding policy about the proprietary nature of safety and effectiveness data.

BIO also believes that the Sponsor, having the most knowledge about the data and their use, should have the opportunity to be involved in determining how to mask data, to review the data post-masking, and to ensure the dataset documentation is appropriate on a case-by-case basis. Any criteria and assumptions used to 'pool' data must be defined for each data set, and the Agency should collaborate with an independent body of experts from both industry and academia when developing pooled data sets. Additionally, data used to develop and populate a database should be curated based on standard data field language/parameters and maintained by a third party (*e.g.*, the Agency or an identified consortium). Furthermore, the Agency, in collaboration with experts from both industry and academia, should develop and articulate plans for database curation, integrity, and sustainability. The Agency should publish for comment the complete process and methods for masking data, which should ensure the opportunity for Sponsors to have input into the masking process on a case-by-case basis, before any data sharing arrangements are implemented.

BIO requests that FDA consider taking the following specific steps when defining the process for masking study data:

- Specify who is responsible for creating the masked data and corresponding data set documentation;
- Clarify timing of when and why masked data will be available (*i.e.*, what is the triggering mechanism?);
- Specify whether masking of the data will be performed for, and tailored to, each research proposal, or will be done once in a generic manner (*i.e.*, irrespective of the particulars of any specific research proposal);
- Specify the format envisioned for the datasets (*e.g.*, CDISC-compliant structure);
- Explain how the Sponsor will be involved in the process, as well as mechanisms for, and extent of, Sponsor notification prior to actual release of data;
- Describe the content and structure of the data sets, as well as the context and limitations under which the data were collected, to facilitate proper analysis and interpretation;
- Separate efficacy data from safety data during masking, as coupling these two factors could lead to unmasking of certain products;
- Require pooling of data for multiple products in a class (*i.e.*, no data sharing or evaluation of a single product, especially with regard to the benefit-risk ratio of individual studies or products);



- Share only representative samples of data, rather than entire data sets from individual trials;
- Mask the identity and key physio-chemical properties of the compound, in order to maintain confidentiality of the compound;
- Consider the types of control groups used in clinical trials and how the control arm could change over time, which would impact the ability to mask the products being used in the study (e.g., an investigational product in an older study could become the active control arm in a newer study conducted after the investigational product is approved);
- Consider the types of measurements included in the dataset, as either newer or obsolete measures could indicate the specific window when a trial was performed (e.g., pooled datasets should be comprised of contemporaneous studies for comparability to reduce likelihood of unmasking portions of a pooled dataset);
- Describe sampling methodology for all random samples, as sampling could bias analyses and may not be sufficient to address specific research questions.

While successfully masking data is an essential element of the strategy for maintaining confidentiality of commercial information, researchers using masked data must also acknowledge its limitations. The accuracy of their analyses may be confounded by masking, and FDA must take this factor into consideration when evaluating qualified research requests.

(2) What limitations, if any, should there be on the Agency's ability to make available the masked data as described previously?

As discussed above, masking safety and effectiveness data may not alter their status as confidential commercial information or trade secrets protected from public disclosure under federal law. In specific cases, release of masked data may cause substantial competitive harm to the submitting Sponsor, and if the masking techniques are insufficient, data disclosure could be even more harmful to the Sponsor's competitive interests. Simply because data are masked and cannot be linked to a particular Sponsor does not mean that they are no longer of the type held in strict confidence by Sponsors, are no longer valuable, or are no longer the result of either innovation or substantial effort.

FDA must not make available trade secret information or confidential commercial information contained in product applications. Examples include but are not limited to:

- Any personal data, including names of Principal Investigators, Data Monitoring Committee or Steering Committee members, or Testing Facilities' personnel;
- Data concerning product formulation, manufacturing, specific study design, and timelines;



- Strategies used by Sponsors to overcome barriers in the development process;
- Information concerning licensing agreements or collaborations with third parties;
- Information identifying suppliers or testing facilities, unless such information has already been publicly disclosed by the Sponsor.

In addition, research efforts on masked data should not be used to influence pending regulatory decisions for NDA/BLAs, and datasets in an open NDA/BLA should not be eligible for release. Finally, FDA should restrict data release for a drug class in which an insufficient number of approved drugs exists for an indication (or for a certain age group within an indication), as it would be impossible to mask the identity of the drug.

(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?

The Health Insurance Portability and Accountability Act (HIPAA) defines de-identified data as "...data that [do] not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual."^{9,10}

⁹ 45 CFR 164.514(a)

¹⁰ 45 CFR 164.514(b)(2) defines the following elements that make up these data:

- Name;
- All geographical subdivisions smaller than a State, including street address, city, county, precinct, zip code, and their equivalent geocodes, except for the initial three digits of a zip code, if according to the current publicly available data from the Bureau of the Census: (1) The geographic unit formed by combining all zip codes with the same three initial digits contains more than 20,000 people; and (2) The initial three digits of a zip code for all such geographic units containing 20,000 or fewer people is changed to 000;
- All elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death; and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older;
- Phone numbers;
- Fax numbers;
- Electronic mail addresses;
- Social Security numbers;
- Medical record numbers;
- Health plan beneficiary numbers;
- Account numbers;
- Certificate/license numbers;
- Vehicle identifiers and serial numbers, including license plate numbers;
- Device identifiers and serial numbers;
- Web Universal Resource Locators (URLs);
- Internet Protocol (IP) address numbers;
- Biometric identifiers, including finger and voice prints;
- Full face photographic images and any comparable images;



In addition to the specific data outlined in 45 C.F.R. 164.514(b)(2), FDA should also consider the application of the second method of de-identification, that of statistical de-identification.¹¹ While this is a general recommendation, it is particularly relevant when patient populations are small and patients may be more readily identified with clinical trials conducted in that disease. Therefore, extra precautions are needed for studies in small patient populations or rare diseases. Additionally, as studies are increasingly conducted globally, FDA should carefully consider relevant regional privacy regulations designed to protect patients participating in clinical trials outside of the United States.

Finally, any use of the data for purposes not contemplated in the initial informed consent process would legally require contacting each human subject and seeking their informed consent. Research institutions do not generally permit general advance authorization by human subjects of unspecified research. In order to seek consent, the subjects would have to be identified and the cost of individually identifying and contacting each patient could far exceed the benefit of the research. For this reason, the use of the data may have to be limited to the initial authorized purpose and any FDA regulations will need to be flexible to contemplate these limitations.

(4) Would regulatory changes facilitate implementation of such a proposal, and if so, what changes would be most useful?

FDA must define a process for how qualified research requests will be evaluated and which data sets will be shared. This activity should *not* be done within FDA review divisions. Establishing an Advisory Committee for data sharing that includes industry representatives may provide a model, satisfying the need for transparency of process and allowing Sponsors input into the conditions of sharing on a case-by-case basis. The governance process should:

- Set forth requirements that research requests are to be submitted in a proposed study protocol that outlines the specific request, data elements, format and planned analysis;
- Specify how the research proposals will be reviewed and determined appropriate to support, including an evaluation of whether the data can support the research objective;
- Specify how FDA will secure the data and prevent unintended release of the full data or masked, de-identified data;

-
- Any other unique identifying number, characteristic, or code (note this does not mean the unique code assigned by the investigator to code the data).

¹¹ 45 C.F.R. 164.514(b)(1) permits disclosure of health information provided that a qualified statistical or scientific expert concludes, through the use of accepted analytic techniques, that the risk the information could be used alone, or in combination with other reasonably available information, to identify the subject is very small.



- Specify how qualified researchers will become aware of data that could address their research objectives;
- Ensure a qualified statistician will lead the analyses for the research project and will be responsible for the integrity and validity of the analyses;
- Clarify who is responsible for distributing the data;
- Ensure that Sponsors are notified about the research request and have an opportunity to provide feedback to FDA;
- Require the qualified researcher to submit for peer-reviewed publication the findings, along with a discussion of the potential limitations and biases;
- Prevent any misuse of data and ensure Agency review prior to submission for potential publications.

Additionally, if researchers “discover” important safety/efficacy findings, FDA will need to determine how the discovery could migrate from meta-analysis of masked data to regulatory actionable status (*i.e.*, impact on labeling), and whether this is even possible if products are not specifically identified (*e.g.*, some members of a class might be better or worse).

(5) Which situations do you believe disclosing masked data would be most useful to advance public health?

We note that the FDA proposal relates to “the ability to make available de-identified and masked clinical and preclinical data derived from marketing applications” (emphasis added). We believe that unique considerations may exist regarding the disclosure of these differing types of datasets, and we would like to further discuss with the Agency the relative merits of disclosing nonclinical data.

Using nonclinical safety data generated from both *in vivo* and *in vitro* surrogate models to accurately identify potential hazards and/or predict clinical outcomes (*i.e.*, translational relevance) is paramount for the refinement of nonclinical models to advance the safest drugs. Furthermore, such refinement enables the ethical use of animals by employing the principles of continued refinement, reduction, and replacement.¹² To this end, meta-analyses of the translational outcomes of pooled, masked, nonclinical safety data and de-identified clinical safety data may be used to understand the limitations and appropriate uses of current preclinical models, as well as to define focused areas for their refinement.¹³ In addition, epidemiological evaluation of

¹² S6 Addendum to Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals (2012) <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM194490.pdf>

¹³ Pierson JB, et al (2013) A public-private consortium advances cardiac safety evaluation: achievements of the HESI Cardiac Safety Technical Committee. *J Pharmacol Toxicol Methods*. doi:pii: S1056-8719(13)00231-1.



de-identified clinical trial safety data and associated risk factors for specific adverse effect classes that are not effectively identified or predicted in current nonclinical models (e.g., idiosyncratic drug-induced liver injury) can be used to refine or develop new preclinical surrogate assessments.^{14,15,16}

In addition, fully contingent upon FDA's ability to successfully address the concerns enumerated above, BIO believes that responsible data sharing may impact the following areas of drug development:

- Investigating orphan indications or other disease states where there are small populations that may benefit from pooling data across multiple studies;
- Establishing surrogate endpoints and developing biomarkers;
- Understanding the course of disease in the untreated arm to help identify better treatment intervention and endpoints;
- Modeling disease outcomes using multivariate approaches;
- Predicting patient populations more or less likely to respond to treatment, and identifying specific circumstances likely to impact response.

CONCLUSION:

BIO appreciates this opportunity to comment on the "Availability of Masked and De-identified Non-Summary Safety and Efficacy Data." We would be pleased to provide further input or clarification of our comments, as needed.

Sincerely,

/S/

Sandra J.P. Dennis
Deputy General Counsel for Healthcare
Biotechnology Industry Organization

/S/

Andrew W. Womack, Ph.D.
Director, Science & Regulatory Affairs
Biotechnology Industry Organization

¹⁴ Lammert C, et al. (2008) Relationship between daily dose of oral medications and idiosyncratic drug-induced liver injury:search for signals. *Hepatology*. 47(6):2003-2009.

¹⁵ Lucena MI, et al. (2009) Phenotypic characterization of idiosyncratic drug-induced liver injury: the influence of age and sex. *Hepatology*. 49(6):2001-2009.

¹⁶ Lammert C, et al. (2010) Oral medications with significant hepatic metabolism at higher risk for hepatic adverse events. *Hepatology*. 51(2):615-620.