JOURNAL OF CLINICAL RESEARCH BEST PRACTICES
Vol. 14, No. 11, November 2018 "Happy Trials to You"

Combining Investigator Data from Three U.S. Government Databases

By Ronald Ranauro, Romiya Barry, and Norman M. Goldfarb

There are three primary U.S. government sources of data about clinical research investigators:

- National Library of Medicine: ClinicalTrials.gov
- FDA: Bioresearch Monitoring Information System (BMIS)
- CMS: Open Payments (Sunshine Act)

Unfortunately, these databases provide only incomplete and inconsistent data, and the data that they do provide are not linked to data in the other databases. The FDA is well aware of these problems (FDA, 2017).

Nevertheless, various researchers have used the databases to understand investigator demographics and activities (Glass, 2009; Glass and Akirtava, 2017; Stergiopoulos, Getz and Blazynski, 2018).

This article presents a study conducted by two of the authors (Ranauro and Barry) to determine to what extent it is possible to combine data from all three databases into a single record for each investigator.

The Study

In this study, we matched and merged data across the three databases to form a more complete picture of investigators and their activities. To make the analysis as apples-to-apples-to-apples as possible, we limited the data in the analysis as follows:

- For BMIS, we included any investigator with a U.S address.
- For Open Payments and ClinicalTrials.gov, we include investigators with a U.S. address and a link to an "interventional" study type.
- To be included in the analysis, an investigator must appear in at least one of the three sources in calendar year 2017:
  - For ClinicalTrials.gov, we included investigators with a trial "start date" of 2017.
  - For BMIS, we included investigators with a "date received" of 2017.
  - For Open Payments, we included investigators with a clinical trial starting in 2017, evidence of payments in 2017, or a 2017 first appearance of the investigator, as reported in the June 2018 data file.

These restrictions are somewhat arbitrary since, for example, BMIS might have received data (in a form FDA 1572) in 2017 for a study that started in 2016. As a result, more records can be matched over a longer time period.

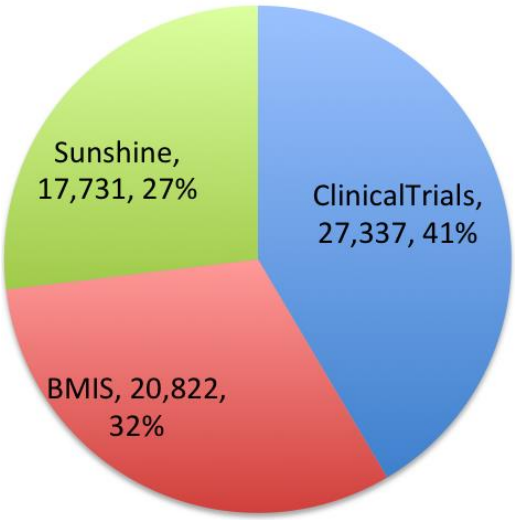
While the BMIS and Open Payments maintain databases of investigators, ClinicalTrials.gov is a database of clinical trials that references investigators. Of the 24,895 trials in ClinicalTrials.gov with start dates in 2017, 19,490 were interventional and, of those, 7,228 had one or more participating investigators with a U.S. address, yielding 27,337 investigator mentions.

Findings

Finding 1. In the three databases, there were 65,890 investigators in 2017 (not counting overlaps).

Figure 1 shows the number of investigator records within each source for 2017, with a total of 65,890:

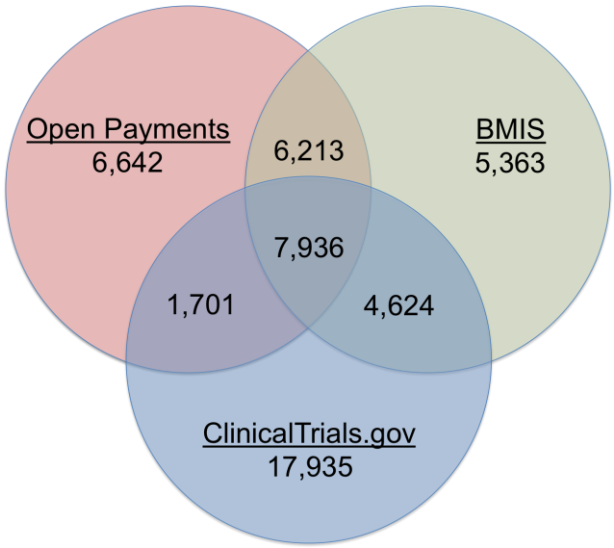
Figure 1. 2017 U.S. Investigators, Interventional Trials, No De-Duplication Across the Three Sources



Finding 2. Investigators overlap incompletely across the three databases

By merging and matching data from the three databases, we identified 50,414 apparently unique investigators. As shown in Figure 2, only 7,936 (15.7%) investigators could be found in all three databases for 2017, 12,564 (24.9%) could be found in two databases, and 29,940 (60.0%) could be found in only one database.

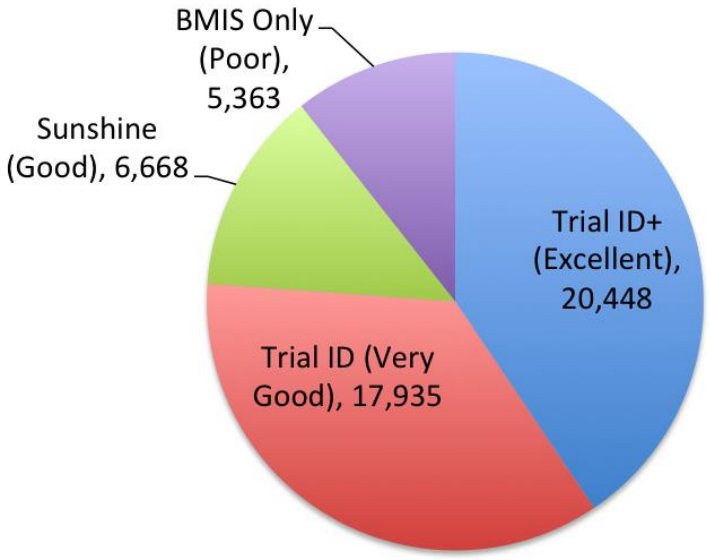
Figure 2. Venn Diagram Showing Investigator Overlaps Across the Three Databases (Not to Scale)



Finding 3. Not all investigators can be matched with clinical trial identifiers.

As shown in Figure 3, 6,668 (13.2%) investigator records have sponsor, payment and therapeutic specialty data, 38,383 (76.1%) can be matched with one or more clinical trials, and 5,363 (10.6%) have only address and date of FDA filing data.

Figure 3. Investigators with Attached Trial Identifiers



Finding 4. Of the investigators that first appear in 2017, 29% first appear in that year.

Finding 5. Some of the investigators that first appeareED in 2017 then became very active.

Figure 4 shows the total number of trial identifiers (of all types) associated with those U.S. investigators connected to any interventional trial with a start date in 2017.

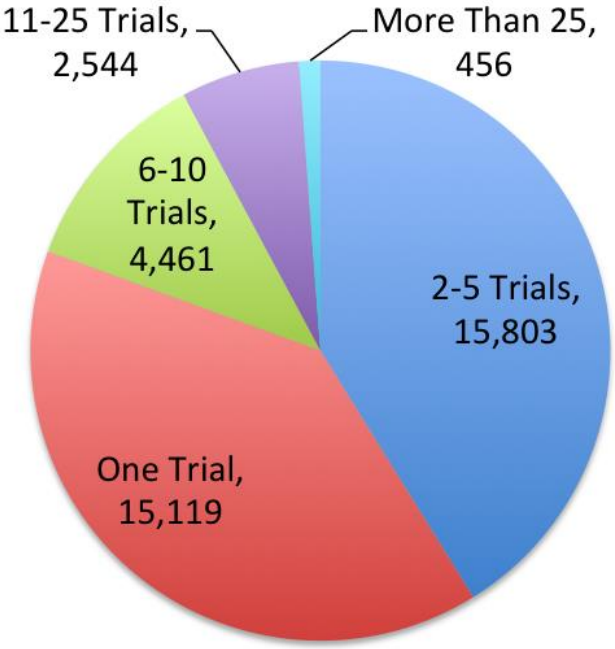
This figure shows that 15,119 (29.4%) of the investigators with one or more attached 2017 interventional trial identifiers appear in either ClinicalTrials.gov or Sunshine for the first time in 2017.

The average investigator with a presence in 2017 has participated in 4.2 trials (mean) and 2.0 trials (median) of any type.

For this analysis, we included trials of all types with start dates as far back as 2000 and as far forward as 2018. Nearly 80% of trials of all types counted have start dates between 2008 and 2018.

Separately, we analyzed the 5,363 BMIS investigator records without a match to the two other databases and found 1,674 (31%) that appear for the first time in BMIS in 2017.

Figure 4. Investigators First Seen in 2017



Conclusions and Recommendations

Given the current limitations of the three databases taken individually, it is impossible to accurately count the number of investigators active in a given year, much less determine with any accuracy attributes like their therapeutic specialties. However, by matching and merging records across the three databases, we can assemble a more comprehensive record for many investigators. Yet gaps remain — for investigator records lacking clinical trial identifiers, there is no way to determine the details of the investigators' research activity.

With coordination among the data sources, additional uses for the data emerge, such as accurately estimating longitudinal trends over time; however, such analyses too are prone to distortion due to exogenous developments like changes in international regulatory policies.

Not only can multiple investigators share the same name, but investigator records sometimes use different versions of an investigator's name. It would be much easier to understand investigator demographics and activity if the following changes were made in the databases:

- Require study sponsors to submit Form FDA 1572s to BMIS with unique clinical trial identifiers.
- At minimum, BMIS and Open Payments should assign a unique identification number to each investigator and use that number consistently. Ideally, this number would be used in all three databases.
- Assign a unique number to each clinical research site, since these names are often entered inconsistently.
- Require use of a standard list of city names.

References

FDA (2017). "Notice to Industry: Bioresearch Monitoring Information System (BMIS) Data." Retrieved from https://www.fda.gov/Drugs/InformationOnDrugs/ucm135162.htm.

Glass, H. E. (2009). "The profile of the one-time 1572 investigator." Contemporary Clinical Trials, 30: 34-39

Glass, H. E., and Akirtava, A. (2017). "What is the Actual Number of Active US Clinical Trial Investigators? Applied Clinical Trials." Retrieved from http://www.appliedclinicaltrialsonline.com/what-actual-number-active-us-clinical-trial-investigators.

Stergiopoulos, S., Getz, K. A., and Blazynski, C. (2018). "Evaluating the Completeness of ClinicalTrials.gov." Therapeutic Innovation & *Regulatory Science*, July:1-11. https://doi.org/10.1177/2168479018782885

Authors

Ronald Ranauro is Founder of Incite Advisors, Inc. Contact him at 1.508.254.8349 or ron@inciteadvisors.com.

Romiya Barry is a Clinical Operations Consultant. Contact her at 1.508.470.1450 or romiya.barry@gmail.com.

Norman M. Goldfarb is Chairman of MAGI. Contact him at 1.650.465.0119 or ngoldfarb@magiworld.org.