



Published in final edited form as:

*Drug Inf J.* 2011 January 1; 45(1): 55–64.

## A New Mechanism for Tracking Publicly Available Study Volunteer Demographics

**Rachael Zuckerman, BA[Research Analyst],**

Tufts Center for the Study of Drug Development, Tufts University, Boston, Massachusetts

**Kenneth Getz, MBA[Senior Research Fellow], and**

Tufts Center for the Study of Drug Development, Tufts University, Boston, Massachusetts

**Kenneth Kaitin, PhD[Director and Professor]**

Tufts Center for the Study of Drug Development, Tufts University, Boston, Massachusetts

### Abstract

The importance of gathering and monitoring aggregate demographic data on the annual population of study volunteers in FDA-regulated clinical trials is widely acknowledged. To date, no formal mechanism exists to capture this information. The Tufts Center for the Study of Drug Development identified and tested a publicly available source of information on clinical trial participant data, NDA Reviews stored in the FDA's drugs@FDA database, to determine its accuracy, reliability, and feasibility. Thirty-seven new drug applications approved between 2006 and 2008 were evaluated and compared with published sources of demographic data. The authors conclude that the approach described here—NDA review extraction—provides reasonably reliable and conservative estimates of study volunteer demographics and can serve as a useful baseline until Clinicaltrials.gov or other, more complete, public sources become available.

### Keywords

Clinical trial; FDA; NDA; NIH; Demographic data

## INTRODUCTION

The importance of tracking and monitoring study participant demographics in clinical research has been widely discussed and acknowledged over the past decade (1–4). Such metrics, if gathered routinely, would provide valuable information to professionals and policymakers. Accurate data would allow users to characterize the clinical research enterprise and to track minority, sex, and special population inclusion. This information would increase understanding of patient access and recruitment effectiveness. Despite ongoing discussion and broad support, no reliable and validated mechanism exists to track industry-funded clinical research.

It is worth noting that the Food and Drug Administration (FDA) itself does not report the number of volunteers in a given year, despite the fact that the agency oversees more than 4,000 active phase 1–3 clinical studies conducted by 27,000 principal investigators annually,

---

Correspondence Address: Kenneth Getz, Tufts Center for the Study of Drug Development, 75 Kneeland Street, Suite 1100, Boston, MA 02111, kenneth.getz@tufts.edu.

Rachael Zuckerman and Kenneth Getz report no relevant relationships to disclose.

No other potential conflict of interest relevant to this article is reported.

at a cost of \$11 billion in study grants (5). This void has not been filled by any known mechanism.

Many fundamental questions regarding FDA-regulated clinical trials cannot be answered easily or reliably at the present time. For example, how many study volunteers in cancer clinical trials this past year are female? Has the number of minority participants in diabetes clinical trials increased over the past year? Are study volunteers who participated in clinical trials that target Alzheimer disease representative of the distributive prevalence of that disease? Whereas it is currently possible to answer these questions for National Institutes for Health (NIH)-regulated clinical trials, this is not the case for FDA-regulated clinical trials.

In 1994, the NIH issued its first version of “Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research” (6). Since that time, the NIH has successfully monitored its study volunteer demographics. In the process and by design, it has reduced minority and gender disparities among study volunteers participating in NIH-funded clinical research programs (7). In 2001, the NIH revised its guidelines and began to withhold funding from researchers under certain circumstances. Researchers are now required to include women and minority patients in prospective phase 3 clinical investigations and to conduct appropriate subgroup analyses to identify differences in outcomes (7).

In 1998, the FDA echoed the NIH’s first guidelines by releasing the “Guideline for the Format and Content of the Clinical and Statistical Sections of an Application.” This guideline suggested that “a table of all investigations pertinent to safety” include the “age range of patients in each study and sex/race distribution” (8). The FDA guideline, however, is not binding. The binding document, *Code of Federal Regulations* “Applications for FDA Approval to Market a New Drug,” is less stringent, suggesting—but not requiring—that both safety and efficacy data be “presented by gender, age, and racial subgroups” (9).

In 2001, a General Accounting Office (GAO) evaluation of demographic data reported to the FDA that underrepresentation of women was no longer an issue (10). However, the GAO did note many other deficiencies. Looking at approved New Drug Applications (NDAs) for New Molecular Entities (NMEs) between 1998 and 2000, the GAO found that women made up more than half of the total participants for each of those NDAs that reported gender data. However, about a third of the applications did not report gender at all, a problem that the GAO felt the FDA still needed to address (10).

Several recent studies have demonstrated that the representation of women in clinical trials has improved since the implementation of the NIH and FDA guidelines. In one of the first studies examining female representation in clinical trials, the FDA examined Biologic License Applications (BLAs) approved between 1995 and 1999 and found gender information for only 14% of blood products, 55% of therapeutics, and 63% of vaccine products. In those cases for which information was available, women made up less than half of the study participants (11). In contrast, an analysis of clinical trials approved between 2000 and 2002 found that women were well represented in general, although earlier-phase trials and some therapeutic areas still predominantly used male subjects (12). Another study over the same time period found that sex was reported for 97% of subjects for approved NMEs between 2000 and 2002. For products indicated for both sexes, the proportion of male and female subjects was nearly equal (13). However, participants’ race was still reported less frequently than sex. In a study of NMEs approved between 1995 and 1999, race could be determined for only 53% of participants (14).

In the absence of a reliable mechanism to monitor patient participation in FDA-regulated, industry-funded clinical research studies, Tufts Center for the Study of Drug Development (CSDD) set out to evaluate one possible approach using publicly available data on clinical

trials of new medical treatments approved by the FDA. A reliable, validated approach would serve as both an initial reporting mechanism and a baseline to evaluate future mechanisms. This study determined whether the method we termed NDA review extraction is a feasible and accurate approach for aggregating demographic information of preapproval subjects for approved products.

## MATERIALS AND METHODS

To examine robust demographic information on clinical trial participants, Tufts CSDD turned to the only known comprehensive source of publicly available data of which we are aware: the FDA's NDA review database. This study method must rely on a public source to ensure that demographic information can be compared over time; however, the NDA often contains partial and incomplete demographic information. We looked at the availability and quality of demographic data in NDA reviews to determine whether they are a feasible way to track study demographics.

There were three main components of our evaluation of the NDA review extraction method. First, we conducted a preliminary analysis to identify the types of demographic information provided in the NDA reviews. Then, we collected available demographic information from NDA reviews using the NDA review extraction method. Finally, we compared our data to other published sources to validate the method.

For the purposes of our analysis, we used data from NDA reviews extracted from the FDA's drugs@FDA database, which contains information on all drugs approved by the Center for Drug Evaluation and Research (15). In total, data from 37 complete NDAs approved between 2006 and 2008 were evaluated using the NDA review extraction method for the completeness of their demographic information.

A preliminary analysis examined one NDA from each year between 2000 and 2008 ( $n = 9$ ) to establish whether demographic data were accessible and consistent from year to year. We selected the first complete NDA review available from that year, looking for all available data on subjects' race, sex, ethnicity, and age for all clinical trials (phases 1–4) presented in the NDA. We also collected background data on each clinical trial, including the trial phase, total number of participants, number of participants in each treatment arm, and whether there was compassionate use of the therapy after the clinical trial was completed. In this preliminary analysis, all data were compiled using the terminology, categories, and units in which they were presented in the NDA.

Following this preliminary analysis, we collected data from 37 NDAs. Tufts CSDD's NDA review extraction method involves collecting study volunteer demographic data by searching the clinical and statistical review sections of each NDA. For the 37 NDAs analyzed, we looked for the phase, overall number of participants, and total number of participants receiving each treatment. Demographic data included participants' sex, race or ethnicity, and the mean, minimum, and maximum age for each trial. Race and ethnicity categories were white, black, Hispanic, Asian, and other, developed from the most common categories found in the preliminary analysis. The NIH guidelines state that race and ethnicity data should be reported separately (3), but we found that these guidelines were not followed often enough for us to collect them as distinct variables. Given the wide variability in the data reported, race and ethnicity categories were not mutually exclusive for all trials. In many cases, only partial demographic data were reported. Gender analyses excluded any programs for sex-specific products.

To test the accuracy and reliability of the NDA review extraction method, demographic information was taken from FDA reviews of NDAs for products approved in 2007. We

included all first approvals for therapeutic products classified by the FDA as chemical types 1–4. First approvals were defined as the first FDA-approved product with a given trade name. Diagnostic products were excluded from this analysis, as there are no comparative data to evaluate the reliability of participant demographic data. For all products included in the analysis, the NDA medical and statistical review sections were examined for demographic information of all reported phase 2, 3, and 4 trials. Additionally, data were collected for five drugs each for 2006 and 2008 as a comparison; these comparison drugs were randomly selected from the approved NDAs in each year meeting our inclusion criteria. To check whether this method missed any trials containing demographic data, five drug applications from 2007 were converted into text documents with optical character recognition using Adobe Acrobat 8 Professional (Adobe Systems Inc, San Jose, CA), and then checked for any clinical trials missed in the initial application review. These applications were searched for the words *demographic*, *white*, *Caucasian*, *black*, and *African* to identify any missed data.

To evaluate the feasibility of this approach, we tracked the amount of time it took to gather data from each review during both the preliminary analysis and using the NDA review extraction method.

All of the demographic data collected were then compared with published data from the *Parexel Statistical Sourcebook*, an independent reference resource containing summary data provided by pharmaceutical and biotechnology company reports, and business and scholarly studies (16).

## RESULTS

### PRELIMINARY ANALYSIS

The preliminary analysis collected demographic information in the form that it was presented in each NDA. After aggregating these data, we were able to revise the NDA review extraction data collection instrument and to define the categories to collect demographic information. The NDA review extraction method collects demographic data on sex, race and ethnicity, and age and trial information on trial phase, total number of participants, and number of participants in each treatment group. The most common race and ethnicity categories were white (80% of NDAs), black (80%), Hispanic (50%), Asian (60%), and other (80%) or equivalent terms. Two of the oldest NDA reviews did not provide any race or ethnicity data. Other data collected for the preliminary analysis were not supplied in a sufficient number of the NDA reviews to be included in the later data collection.

For the preliminary analysis, data collection took 60 to 90 minutes for each review. Although time intensive, our preliminary analysis confirmed the feasibility of a larger evaluation effort. To that end, data from phases 2, 3, and 4 trials were gathered for 3 years of NDA approvals (2006–2008) in the larger evaluation study.

### ASSESSING DATA QUALITY

Some demographic data were available for all products, with the most complete data available on participants' sex and/or demographic data for pivotal phase 3 trials. The availability and extent of data on race or ethnicity, age, and demographics for additional phase 2 and 3 trials varied by review.

Twenty-seven drugs were included in the analysis of the 2007 approvals, and five drugs each from 2006 and 2008 approvals were used for comparison (see Table 1). Twenty-nine of the total 37 drugs evaluated for this analysis were subsequently approved for adults only;

seven were approved for some pediatric populations as well as adults; and one was approved only for a pediatric indication. Nine drugs had orphan designations.

Tables listing clinical studies, required in an NDA, were always found in the same section of the review. Demographic information was typically found either in the medical review's integrated review of efficacy or appendix and was located in the same place for each trial within an application. These data were most often in tabular form, but not always. Some applications had demographic data in the statistical review that were not duplicated in the medical review. Many products had multiple review cycles. In this case, trial demographics were only presented in the first submission where the trial was mentioned.

Not all NDA reviews provided demographic data for all studies referenced in the NDA. For example, the clinical review of nebivolol contains a list of all supported studies and their results, but does not contain demographic data for any of the 62 studies that do not report efficacy results. This is similar to the review for lanreotide, which did not contain efficacy results or demographic data for 42 safety and efficacy studies using patient subjects.

By limiting the amount of data and focusing on specific application sections from the preliminary analysis, we were able to decrease the amount of time it took to gather and input data from each NDA. Some reviews of NDAs were as short as 45 minutes, while others took us over an hour.

For five of the products approved in 2007, we converted the entire application review into text and used a search function to see if any trials were missed in the NDA review extraction. For one of these products, we found an ongoing trial that was not in the original data set. The study description provides demographic data on sex, age, treatment, and whites enrolled in the study but no additional race or ethnicity data. This study has subsequently been added to the analysis presented in this study.

## ASSESSING COMPLETENESS OF DEMOGRAPHIC DATA

The demographic results derived through the NDA review extraction method are consistent with those published elsewhere (10–12), but the total number of subjects is lower.

To validate the data across the 3 years, we analyzed the data in two ways: by entire drug programs, and by programs stratified by phase. Of these two analyses, the 3 years of data collected were comparable by whole drug program, but the drugs from 2007 were not consistent with those from 2006 and 2008 when stratified by phase (see Table 2). For this reason, all subsequent analyses were done by whole drug programs.

For products approved in 2007, we reviewed an average of four trials with demographic information per product (range 2–11 trials). Only 12 of the 27 products approved in 2007 had demographic information for any phase 2 trials, with two additional product reviews providing data for phase 2/3 trials. For phase 3 trials, 22 applications provided demographic data for at least one trial; on average these applications provided data from two phase 3 trials. Two products did not specify trial phase, and one had no phase 3 trials in the application.

The NDA reviews of the 2007 approvals included an average of 1,680 subjects. We compared this to the *Parexel Statistical Sourcebook*, which publishes data on the total number of subjects for each NDA from product labels and sponsors (16). The NDA review extraction method produced consistently lower numbers of subjects than the total number of subjects reported in the *Parexel Statistical Sourcebook* (Table 3). This was also true when comparing individual products (Table 4). Across the 11 products approved in 2007 that are

contained in both our analyses and in the *Parexel Statistical Sourcebook*, the NDA review extraction method data contain 12% fewer subjects in total.

Demographics of the study populations for the 27 products approved in 2007 are presented in Table 5. Gender was compared for the 25 products indicated for both males and females, all of which provided subject sex information for all trials reviewed. Males comprised a little more than half (54.3%) of all subjects in these trials, consistent with previous studies (4,8). Gender data were provided more often than race and ethnicity data (99.4% and 94.2% of subjects, respectively). Overall, 17 of the 27 products approved in 2007 provided race or ethnicity data for all subjects in all trials reviewed. Most of these subjects were white (82.3%), while a much smaller number were black (7%). Reporting categories for race and ethnicity data also varied between product reviews. Based on the available data, representation of whites and Asians is consistent with US national demographics, although all other groups are underrepresented (Table 5).

## DISCUSSION

The results of this evaluation confirm that Tufts CSDD's NDA review extraction method of gathering data from the clinical and statistical sections of reviews in the drugs@FDA database is feasible and reliable in the near term. The NDA reviews are publicly available for all products approved by the FDA, and they provide, in one place, information on the whole clinical program.

The aggregate number of subjects in the data we collected is 12% lower than the only other published source of such information (16). Some of this difference may be due to the fact that phase 1 subjects are included in data published in the *Parexel Statistical Sourcebook* and omitted in our evaluation of NDA reviews. Use of an adjustment factor may mitigate this discrepancy. In terms of capturing available data, our collection method was generally accurate; we found only one missed trial when rechecking five NDA reviews. The single missed trial among the five applications rechecked accounted for only 1.7% of the total subjects for these products.

Overall, about one quarter of the NDA reviews were missing some sex data, and one third were missing some race data. However, this translates into very few individual subjects; the NDA review extraction method captures demographic data for over 90% of the subjects included. This is an improvement from application reviews in the 1990s, where only 53% of the participants' race could be determined (10).

In our analyses, more demographic data were missing from phase 2 studies than from phase 3 studies. This may be due to the "Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research" (6), which only requires that demographic data be provided for phase 3 studies. Nevertheless, to get an accurate picture of clinical trial participants, sponsors need to go beyond the NIH guidelines and report phase 2 subjects' demographic information as well.

The demographic distribution for those subjects in which information was provided was consistent with previous studies (17–19). Some of the observed disparity in race and ethnicity proportions between clinical trial and census data may be due to different assignment methods. For example, the US Census asks about Hispanic ethnicity separately from race and allows individuals to select multiple races as well as Hispanic ethnicity. Most trials reviewed for this analysis, however, reported race and ethnicity combined, placing individuals in a single category.

It is possible that the US population is not an appropriate study population for some drugs in development. The globalization of clinical trials may mean that sponsors do not want to match their study populations to the US population. Additionally, clinical trial populations are often chosen to represent disease prevalence, which can be different from the general population. If an alternate study population is more appropriate, it is still important to provide the demographics of the subjects studied and may also be necessary to define the demographics of the study population that subjects should be compared to.

Our results follow previously observed minority underrepresentation among clinical research study subjects (6,17,18). Such a continuing trend is a concern for both the generalizability of study results to the intended population and for understanding treatment effects in subpopulations (2).

The largest disadvantage of the NDA review extraction method is the considerable amount of time it takes to gather data. Most applications took over an hour to review. In the future, other sources may provide more comprehensive data in an accessible format, but until such time, NDA review extraction is an adequate method.

## CONCLUSION

The NDA review extraction method is a valuable tool for collecting demographic data on approved products. Our method uses the only publicly available source of comprehensive clinical program information, approved product reviews, to provide highly accurate data of clinical trial subjects' demographics.

In the future, other, less labor-intensive methods may become available to gather comprehensive clinical trial data. The most promising alternative method is contained in the FDA Amendments Act of 2007 (FDAAA), which requires that the results of some trials be published on a publicly available database. The results must include demographic information and should be published on the ClinicalTrials.gov online registry (20). This requirement is only binding for trials submitted to the registry after the passage of FDAAA. Because these regulations went into effect in the fall of 2008, more time is needed to evaluate the registry as a data source for monitoring study volunteer demographics.

At this point, it is unclear whether the ClinicalTrials.gov database will provide more complete demographic data on development programs than the NDA reviews. The only required demographic information for ClinicalTrials.gov is age and gender. Moreover, all demographic measures (age, gender, race, and ethnicity) can be reported in a predefined format or customized, meaning that reporting does not have to be standardized at this time (21). Ideally, an approach to gathering and tracking aggregate national clinical trial volunteer data would be both standardized and routinely used.

If sponsors do report comprehensive demographic data, the publication format of ClinicalTrials.gov is much more conducive to data collection than approved product reviews. Trials are searchable by drug name and demographics are reported in tabular form. As more information is available in the ClinicalTrials.gov database, further study of the accuracy and completeness of these data is needed.

For the time being, we believe that Tufts CSDD's NDA review extraction method is a reasonable approach for tracking and reporting aggregate clinical trial study volunteer demographics.

## Acknowledgments

Kenneth Kaitin is supported in part by grant UL1 RR02572 from the National Center for Research Resources.

## References

1. Powell JH, Fleming Y. Making medicines for America: the case for clinical trial diversity. *J Natl Med Assoc.* 2000; 92:507–514. [PubMed: 11152082]
2. Garber M, Arnold RM. Promoting the participation of minorities in research. *Am J Bioeth.* 2006; 6(3):W14–W20. [PubMed: 16754431]
3. Wendler D, Kington R, Madans J, et al. Are racial and ethnic minorities less willing to participate in health research? *PLoS Med.* 2006; 3(2):e19. [PubMed: 16318411]
4. Sung NS, Crowley WF, Genel M, et al. Central challenges facing the national clinical research enterprise. *J Am Med Assoc.* 2003; 289:1278–1287.
5. Getz K. Sizing up the clinical research market. *Appl Clin Trials.* 2010; 19(3):32–34.
6. Sheet 11—inclusion of women and minorities in study populations: guidance for IRBs and principal investigators. Office of Human Subjects Research; [accessed November 4, 2009]. <http://ohsr.od.nih.gov/info/sheet11.html>
7. NIH policy and guidelines on the inclusion of women and minorities as subjects in clinical research—amended. Office of Extramural Research; October 2001 [accessed August 7, 2009]. [http://grants.nih.gov/grants/funding/women\\_min/guidelines\\_amended\\_10\\_2001.htm](http://grants.nih.gov/grants/funding/women_min/guidelines_amended_10_2001.htm)
8. FDAMA women and minorities working group report. Center for Drug Research and Evaluation; [accessed August 7, 2009]. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm080616.pdf>
9. Content and format of an application, 21 CFR Section 314.50. 1998.
10. Women sufficiently represented in new drug testing, but FDA oversight needs improvement. General Accounting Office; July 2001 [accessed August 7, 2009]. [www.gao.gov/cgi-bin/getrpt?GAO-01-754](http://www.gao.gov/cgi-bin/getrpt?GAO-01-754)
11. Participation of females in clinical trials and gender analysis of data in biologic product applications. US Food and Drug Administration; [accessed August 7, 2009]. <http://www.fda.gov/BiologicsBloodVaccines/DevelopmentApprovalProcess/InvestigationalNewDrugINDorDeviceExemptionIDEDProcess/ucm094300.htm>
12. Taylor, N. More women now in clinical trials; study. *Outsourcing-pharma.* [accessed December 2, 2008]. <http://www.outsourcing-pharma.com/Clinical-Development/More-women-now-in-clinical-trials-study>
13. Yang Y, Carlin AS, Faustino PJ, et al. Participation of women in clinical trials for new drugs approved by the Food and Drug Administration in 2000–2002. *J Womens Health.* 2009; 18:303–310.
14. Evelyn B, Toigo T, Banks D, et al. Participation of racial/ethnic groups in clinical trials and race-related labeling: a review of new molecular entities approved 1995–1999. *J Natl Med Assoc.* 2001; 93(12 Suppl):18S–24S. [PubMed: 11798060]
15. US FDA. *Drugs@FDA.* [accessed November 15, 2010]. <http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm>
16. Mathieu, MP., editor. *PAREXEL's Bio/Pharmaceutical R&D Statistical Sourcebook 2009/2010.* Waltham, MA: PAREXEL International Corporation; 2009. Emerging clinical trial enrollment benchmarks: trial size statistics for new drugs approved in 2008; p. 166-168.
17. Mak WW, Law RW, Alvidrez J, et al. Gender and ethnic diversity in NIMH-funded clinical trials: review of a decade of published research. *Adm Policy Ment Health.* 2007; 34:497–503. [PubMed: 17690976]
18. Heiat A, Gross CP, Krumholz HM. Representation of the elderly, women, and minorities in heart failure clinical trials. *Arch Intern Med.* 2002; 162:1682–1688. [PubMed: 12153370]



19. Getz K, Faden L. Racial disparities among clinical research investigators. *Am J Ther.* 2008; 15(1): 3–11. [PubMed: 18223347]
20. Food and Drug Amendments Act (FDAAA) of 2007. Pub. L. 110-85, 121 Stat. 904 (September 27, 2007).
21. ClinicalTrials.gov. “Basic results” data element definitions. Draft. US Food and Drug Administration; [accessed September 2, 2009]. [http://prsinfo.clinicaltrials.gov/results\\_definitions.html](http://prsinfo.clinicaltrials.gov/results_definitions.html)

TABLE 1

Approved Products Included in Tufts CSDD NDA Review Extraction Tracking Mechanism Evaluation

Product Generic (Trade)	NDA Approval Date	Total Number of Subjects Collected From NDA Review Data	Total Number of Trials Collected From NDA Review Data
Mesalamine (Lialda)	1/16/2007	603	2
Diclofenac epolamine (Flector)	1/31/2007	1,137	4
Cyclobenzaprine hydrochloride (Amrix)	2/7/2007	504	2
Lisdexamfetamine dimesylate (Vyvanse)	3/5/2007	342	2
Aliskiren hemifumarate (Tekturna)	3/13/2007	7,060	7
Lapatinib ditosylate (Tykerb)	3/16/2007	1,030	3
Eculizumab (Soliris)	4/12/2007	184	2
Retapamulin (Altabax)	4/16/2007	727	2
Zoledronic acid (Reclast)	5/9/2007	357	2
Fluticasone furoate (Veramyst)	4/27/2007	4,002	11
Rotigotine (Neupro)	5/30/2007	1,164	3
Formoterol fumarate (Perforist)	5/11/2007	1,006	4
Temsirolimus (Torisel)	6/15/2007	737	2
Ketoconazole (Extina)	6/12/2007	1,781	2
Ambrisentan (Letaris)	8/30/2007	393	2
Amlodipine besylate; valsartan (Exforge)	6/20/2007	5,052	4
Tretinoin (Atralin)	7/27/2007	1,565	4
Maraviroc (Selzentry)	8/6/2007	1,049	2
Lanreotide acetate (Somatuline Depot)	10/12/2007	868	11
Doripenem (Doribax)	10/12/2007	2,238	5
Reltegravir potassium (Isentress)	10/16/2007	667	2
Ixabepilone (Ixempra)	10/29/2007	992	4
Nilotinib hydrochloride monohydrate (Tasigna)	10/30/2007	385	2
Brimonidine tartrate; timolol maleate (Combigan)	11/14/2007	2,645	5
Methoxy polyethylene glycol-epoetin beta (Mircera)	12/13/2007	2,299	6
Sapropterin dihydrochloride (Kuvan)	12/17/2007	712	3
Nebivolol hydrochloride (Bystolic)	2/1/2007	6,021	9
Insulin recombinant human (Exubera)	1/27/2006	4,166	16
Anidulafungin (Eraxis)	2/17/2006	1,019	5
Varenicline tartrate (Chantix)	5/10/2006	5,537	6
Ranibixumab (Lucentis)	6/30/2006	1,323	3
Posaconazole (Noxafil)	9/15/2006	1,202	2
Regadenoson (Lexiscan)	4/10/2008	1,871	2
Eltrimbopag olamine (Promacta)	11/20/2008	231	2
Tapentadol hydrochloride (Tapentadol Hydrochloride)	11/20/2008	2,296	6
Choline fenofibrate (Trilipix)	12/15/2008	2,698	3

<b>Product Generic (Trade)</b>	<b>NDA Approval Date</b>	<b>Total Number of Subjects Collected From NDA Review Data</b>	<b>Total Number of Trials Collected From NDA Review Data</b>
Perixafor (Moxobil)	12/15/2008	647	4

**TABLE 2**

Mean Number of Subjects and Trials Included in the Tufts CSDD NDA Review Extraction Tracking Mechanism Evaluation by Year

Measure	NDA's Approved in 2007 (n = 27 Applications)	NDA's Approved in 2006 (n = 5 Applications)*	2008 (n = 5 Applications)*
Total subjects	1,680	2,649	1,548
Phase 2 trials	218	621	541
Phase 3 trials	1,298	2,253	1,332
Total trials	3.9	6.4	3.4
Phase 2 trials	1.6	2.7	3
Phase 3 trials	3.1	4.6	2.2

\* Five applications were randomly chosen from 2006 and 2008 to validate the data collected in 2007.

**TABLE 3**

Total Subjects per Application as Collected by the Tufts CSDD NDA Review Extraction and Company-reported NME Program Sizes in the *Parexel Statistical Sourcebook*

Data Collected From NDA Reviews (Phase 2–3 Trials)		<i>Parexel Statistical Sourcebook</i> (Phase 1–3 Trials)						
Year	Mean	Median	Range	n*	Mean	Median	Range	n
2006	2,649	1,323	1,019–5,537	5	2,848	2,150	911–6,700	11
2007	1,680	1,006	184–7,060	27	2,239	1,076	404–6,745	11
2008	1,549	1,871	231–2,698	5	2,175	1,838	674–4,826	15

\* Five applications were randomly chosen from 2006 and 2008 to validate the data collected in 2007.

**TABLE 4**

Product Program Size as Collected by the Tufts CSDD NDA Review Extraction and Company-reported NME Program Sizes in the *Parexel Statistical Sourcebook*

Product	Approval Year	NDA Review Extraction (Phase 2–3 Trials)	<i>Parexel Statistical Sourcebook</i> (Phase 1–3 Trials)
Chantix	2006	5,537	5,305
Noxafil	2006	1,323	3,038
Eraxis	2006	645	1,230
Selzentry	2007	1,049	1,076
Vyvanse	2007	342	404
Bystolic	2007	6,021	6,745
Ixempra	2007	992	928
Tekturna	2007	7,060	6,460
Neupro	2007	1,164	1,509
Doribax	2007	2,238	2,117
Isentress	2007	667	920
Altabax	2007	727	3,000
Kuvan	2007	712	747
Torisel	2007	737	726
Mozobil	2008	647	750
Lexiscan	2008	1,871	2,165

**TABLE 5**

Demographics Collected by the Tufts CSDD NDA Review Extraction as Compared to US Census Data

Measure	2007 (n = 27 Applications)		2000 Census Data
	Total	Median	
Subjects	45,363	1,006	
Trials	106	3	
Male	54.3%	52.4%	49.1%
Race or ethnicity			
White	74.1%	75.8%	75.1%
Black	9.6%	7.3%	12.3%
Hispanic	7.6%	3.1%	12.5%
Asian	3.3%	1.5%	3.6%
Other	3.9%	2.4%	8.9%
Treatments*			
Placebo	16.6%	19.9%	
Comparator drug	19.9%	19.6%	
Study drug	62.3%	63.1%	

\* Placebo indicates placebo only; Comparator drug represents all subjects receiving a comparator drug but not the study drug; Study drug indicates all subjects getting at least one dose of the study drug.