



ARTICLE

Assessing the Impact of the NIH CTSA Program on Clinical Trials Registered With ClinicalTrials.gov

Arthur Berg^{1,*}

Since 2006, the US Congress has appropriated ~ 7 billion dollars in total toward the (CTSA) program of the National Institutes of Health (NIH), representing ~ 1.5% of the NIH total budget. There is no doubt this investment has led to substantial improvements in clinical and translational research, but the impact of these large NIH-sponsored awards to academic medical centers have largely been documented by anecdotal accomplishments. This paper provides a purely quantitative assessment of the impact of these awards on clinical trials registered on ClinicalTrials.gov. In particular, we find a dramatic increase in the number of registered clinical trials and clinical trial enrollment associated with the CTSA grant award. Additionally, the impact is shown to be magnified with the number of years of receiving CTSA funding.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

✓ The author of this manuscript, in an article published in this journal in 2013, presented an analysis of the impact of the Clinical and Translational Science Award (CTSA) program on registered clinical trials on ClinicalTrials.gov that utilized a much different methodological approach.³ That paper only identified a statistically significant effect of the CTSA program on registered clinical trials, but it did not provide an explicit assessment of the impact.

WHAT QUESTION DID THIS STUDY ADDRESS?

✓ Does clinical trial registration and enrollment change after an institution receives CTSA funding, and if so by how much?

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

✓ This study provides a quantitative assessment of the impact of the CTSA program on translational medicine in the United States.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

✓ This study shows that the CTSA program has a perceived substantial impact on clinical trial registration and enrollment, which provides more support for this vital National Institutes of Health (NIH) program that seeks to promote translational science in many different ways.

The National Institutes of Health (NIH)-sponsored Clinical and Translational Science Award (CTSA) program began in late 2006 with a consortium of 12 academic medical centers and after 4 years or so the program grew to encompass ~ 60 different institutions. The CTSA consortium provides an integrated home for conducting clinical and translational science and provides resources to researchers to apply the new scientific knowledge and techniques to patient care. The scope of the program is highly interdisciplinary and encourages collaborations with researchers spanning different scientific disciplines. There has been a substantial impact of the CTSA program on clinical and translational research. One fairly comprehensive assessment of the impact is documented in the recent article “Scope, Influence, and Interdisciplinary Collaboration: The Publication Portfolio of the NIH Clinical and Translational Science Awards (CTSA) Program From 2006 Through 2017.”¹

The strategic goals of the CTSA program have slightly changed over the years, but from early on one of those goals has been to “advance translational research to move

basic laboratories and knowledge into clinical testing.”² This paper seeks to assess the impact of this goal by analyzing the number of clinical trials of CTSA member institutions on ClinicalTrials.gov over the past 18 years. In 1997, the US Congress passed the Food and Drug Administration Modernization Act, which is a federal law requiring clinical trials to be registered, and, in 2000, the NIH National Library of Medicine developed ClinicalTrials.gov to facilitate the registry of clinical trials. The ClinicalTrials.gov now contains over 321,000 trials from countries all over the world.

A similar assessment of the CTSA program was taken about 5 years ago in the publication “Assessing the Impact of the NIH CTSA Program on Institutionally Sponsored Clinical Trials.”³ In this earlier analysis, CTSA-funded institutions were matched with other US institutions not receiving CTSA funding and a statistically significant effect of the CTSA funding on clinical trial registration was identified. However, that approach did not provide a good means of quantifying effect in terms of percentage increase of trial registration or participant

¹Department of Public Health Science, Division of Biostatistics & Bioinformatics, Penn State College of Medicine, Hershey, Pennsylvania, USA. *Correspondence: Arthur Berg (berg@psu.edu)



Figure 1 Clinical and Translational Science Award (CTSA) institutions and their funding status from 2006–2019. Int'l, international; NIH, National Institutes of Health.

enrollment nor was it clear how these effects changed over time. Substantially more data are used in the current analysis and incorporates a much different methodological approach that is able to overcome these limitations of the prior analysis.

This paper uses a comprehensive Bayesian model to predict the number registered clinical trials in a given year for a given source/institution. In particular, the following general methodological approach is taken.

1. The number of registered clinical trials and number of participants enrolled in these trials is extracted from the ClinicalTrials.gov database for each institution and each year from 2000 to 2018.
2. Data values for the institutions receiving CTSA funding in the respective years were deleted and replaced with missing values to be predicted.
3. A Bayesian model that incorporates model parameters for each different institution (a total of 344 institutions

was considered), a parameter for each year from 2000 to 2018, and a parameter for each class of institution (CTSA institution, US non-CTSA institution, international institution, NIH institution, and pharmaceutical company).

4. The Bayesian model was used to predict the data values (number of registered clinical trials and number of participants enrolled) for each source and year combination.
5. The predicted values are compared with the actual values to quantify the effect of the CTSA program on the number of registered clinical trials and enrollment numbers.

With this innovative methodological approach, we can attempt to assess the impact of the CTSA program by minimizing the confounding effect of time. This analysis does not presume to avoid all potential confounding factors, but, on

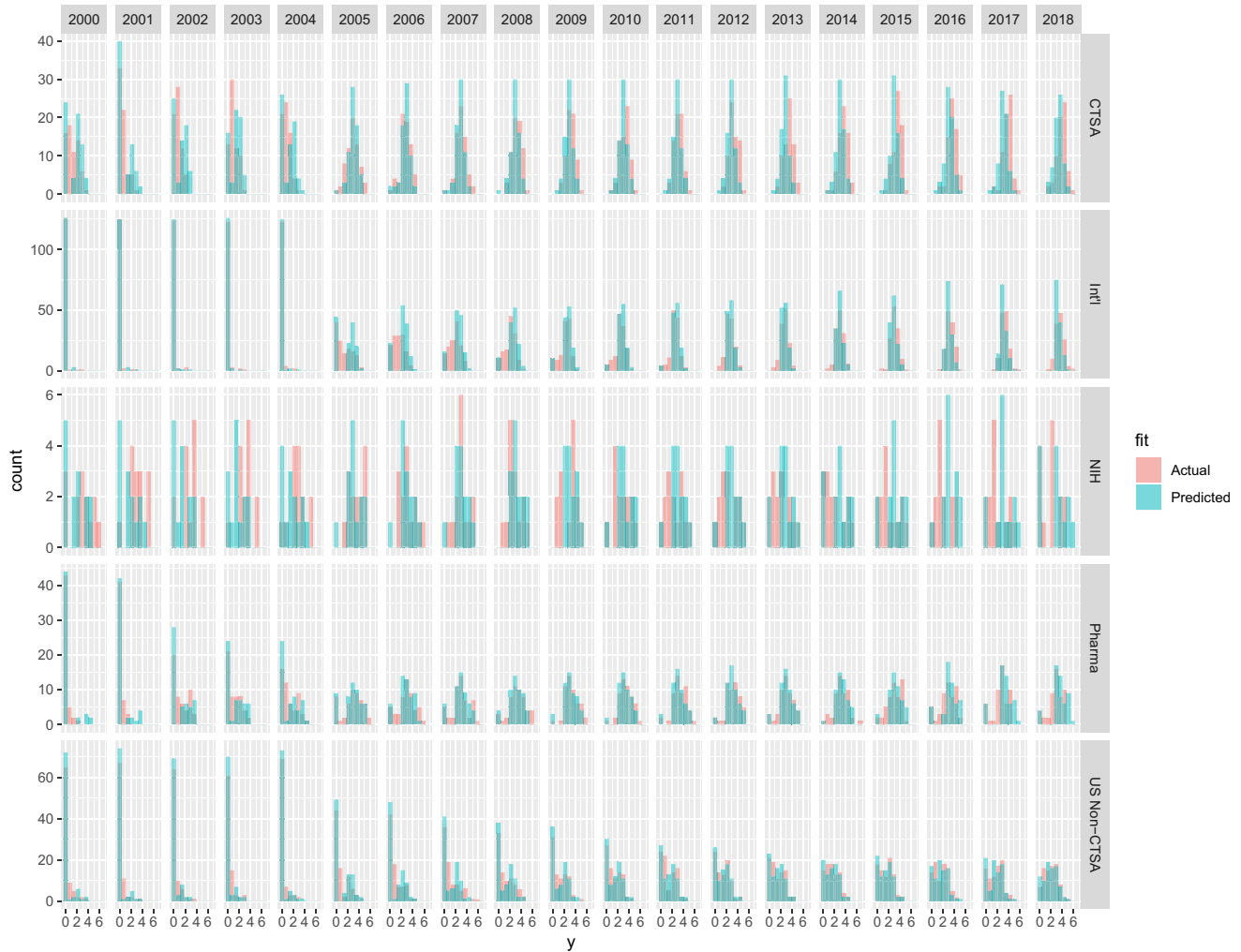


Figure 2 Histograms of number of trials (actual and predicted) by year and group (Clinical and Translational Science Award (CTSA), US non-CTSA, National Institutes of Health (NIH), international (Int'l), and pharma).

the whole, this approach can broadly quantify the effects of the CTSA program-based utilizing the vast information that is incorporated into the model.

METHODS

General approach

Clinical trial registry data were extracted from ClinicalTrials.gov for CTSA institutions as well as other institutions, including non-CTSA academic medical centers in the United States (84 different sources), US Government Institutions (14 different sources), academic medical centers outside the United States (129 different sources), and pharmaceutical companies (51 different sources). The number of clinical trials as well as the total enrollment in clinical trials for each institution for each year from 2000 to 2018 was recorded. In creating a modeling data set, outcome data (number of registered clinical trials and enrollment numbers) were removed for the sources that received a CTSA award for the respective year. A flexible Bayesian model was fit to the data and it was used to predict the outcome data for each source in each year. Although outcome data for the CTSA awardees

were removed, the Bayesian model can reasonably infer the outcome values based on (i) data available from the same source for the years in which there was no CTSA funding, (ii) data available from other sources in the same year, and (iii) data from other sources from other years. Overall, the outcome data increased year-over-year, and the flexible model uses the information from all available sources to incorporate such trends in its predictions. The accuracy of the fitted model can be assessed by how close the predictions match the actual values for all non-CTSA sources. As the model seems to fit the non-CTSA institutions well and without a clear systemic bias, difference between the predictions and the actual values for the CTSA institutions is attributed largely to the CTSA-funding mechanism.

Technical methods

The first step in this analysis was to determine which institutions and in which years received CTSA funding. Institutions receiving CTSA funding (currently referred to as CTSA program hubs) for the years 2013–2018 is available from the Center for Leading Innovation &

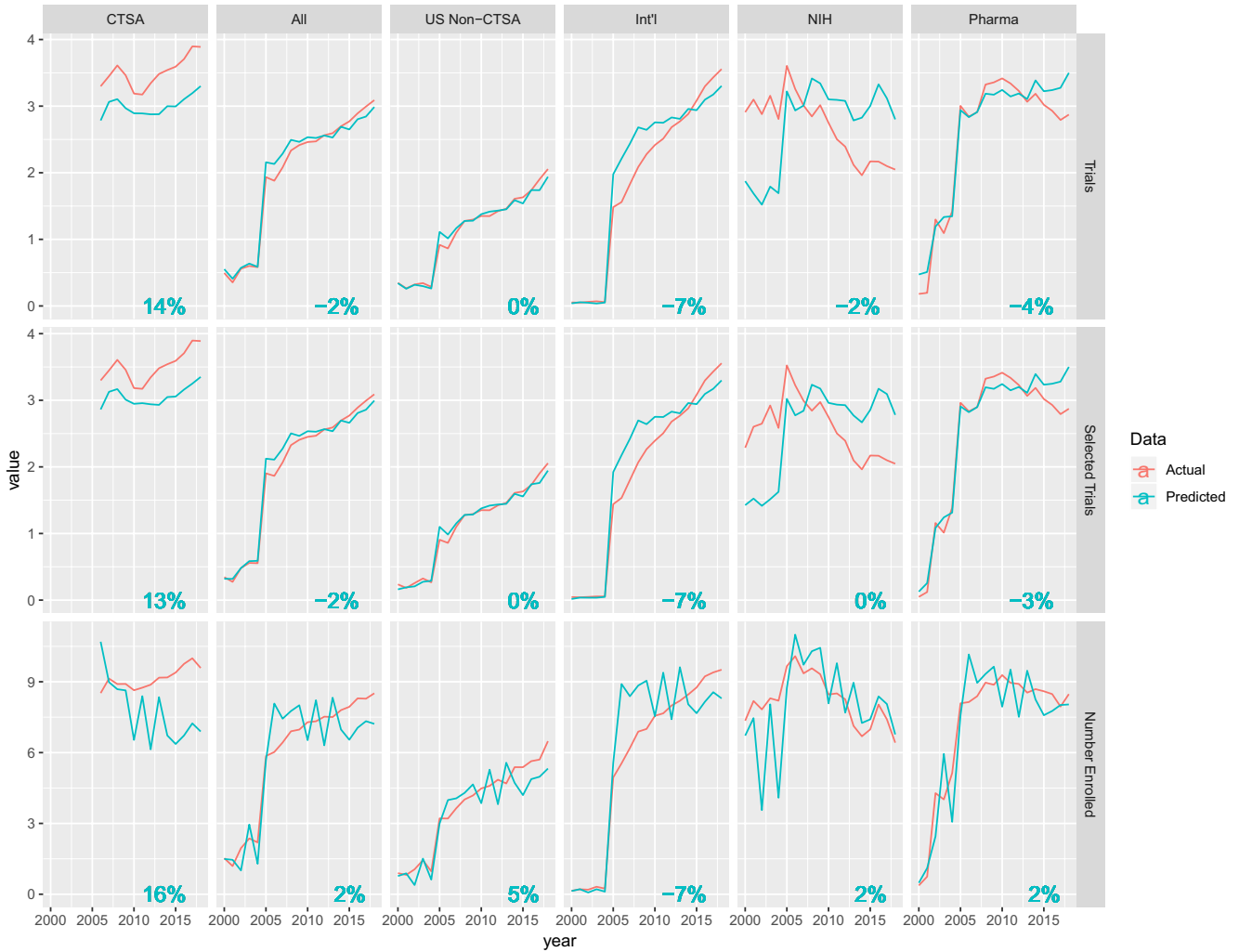


Figure 3 Log-transformed counts of actual and predicted values of the primary measures (number of trials, number of trials with positive enrollment, number of enrolled participants) by year and group (Clinical and Translational Science Award, US non-CTSA, National Institutes of Health, international, and pharma).

Collaboration at <https://clic-ctsa.org/ctsa-program-hub-directory>. Various other archived documents were used to determine the institutions receiving funding from 2006–2012. A summary of all of the years each funded institution received CTSA funding is depicted in **Figure 1**. Once funded, institutions tend to receive renewal funding, although sometimes with 1-year or 2-year lapse in funding between renewals. More rarely, institutions may lose CTSA funding for up to 4 years.

The next step in the data collection involves extracting the ClinicalTrials.gov data. This was done using resources provided by The Clinical Trials Transformation Initiative (CTTI). The CTTI is a public-private partnership that seeks to develop and drive adoption of practices that will increase the quality and efficiency of clinical trials. In particular, the CTTI maintains the Aggregate Content of ClinicalTrials.gov relational database that is refreshed daily. The Aggregate Content of ClinicalTrials.gov includes all of the protocol and results of the data elements for studies that are publicly available at ClinicalTrials.gov.

For this analysis, we utilize just the “studies.txt” table, which contains the variables study ID, study start date, estimated/actual enrollment, and source. The enrollment number may or may not be available and if it is available it may be based on actual enrollment or targeted enrollment. Unfortunately, the data source only provides the actual or targeted enrollment and not both. Based on these available data, three primary measures were used:

1. Number of clinical trials per source per year.
2. Number of clinical trials with positive enrollment per source per year.
3. Total enrollment numbers per source per year.

It is noted that CTSA hubs are often affiliated with multiple sources, such as a source for the university, a source for the affiliated cancer center, a source for the affiliated children’s hospital, separate sources for affiliated hospitals, etc. The specific mapping of the various sources to the CTSA hubs are documented in the **Supplementary Material S1**.

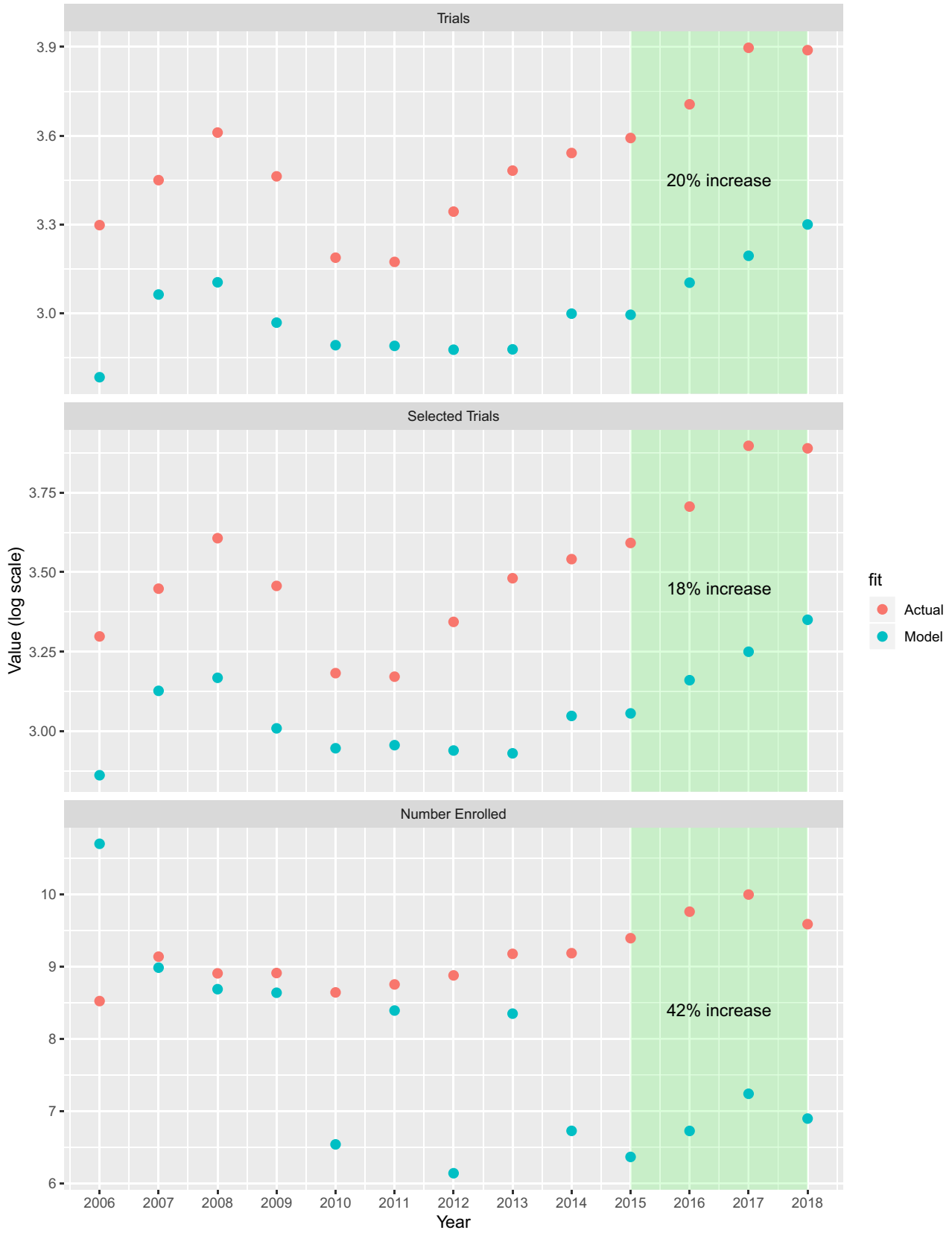


Figure 4 The left-hand column of **Figure 3** is shown in this figure with differences between the actual and predicted outcomes from 2015 to 2018 highlighted. CTSA, Clinical and Translational Science Award.

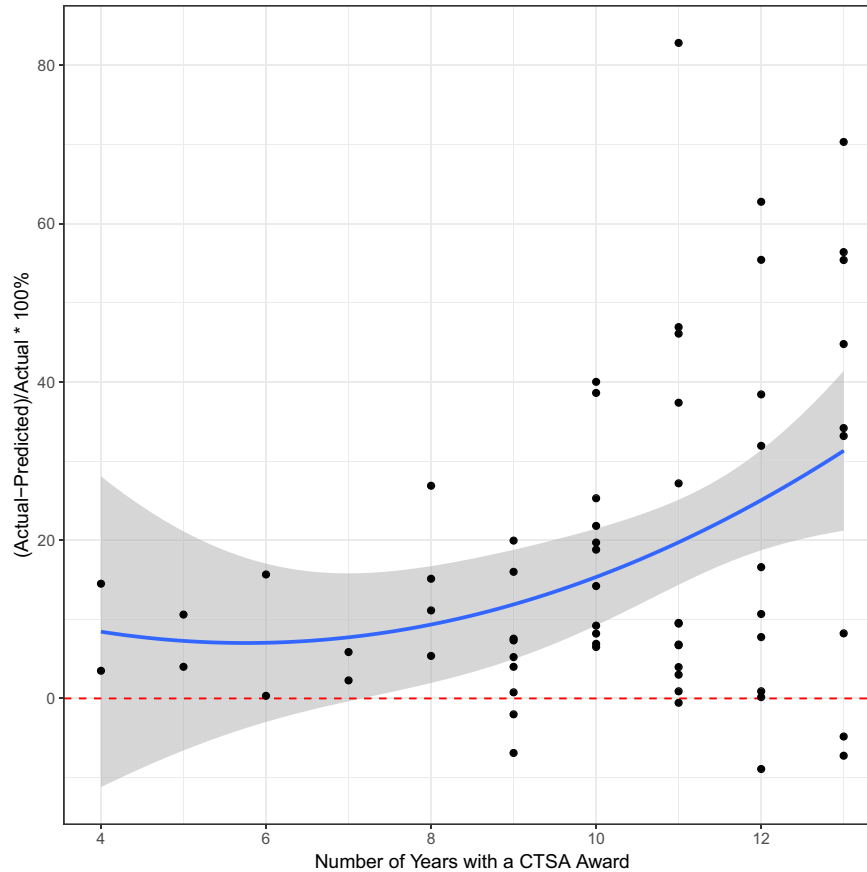


Figure 5 Relative difference of actual and predicted counts by number of years having Clinical and Translational Science Award funding.

In addition to the 66 CTSA hubs, data were collected on 278 other top-contributing clinical trial sources, including 84 US non-CTSA sources, 14 NIH-related sources, 129 international sources, and 51 pharmaceutical sources.

The final data set has 344 sources/institutions * 19 years for a total of 6536 rows and contains the following variables: source/institution, year, source group (CTSA, US non-CTSA, NIH, international, and pharma), the three primary measures (number of clinical trials, number of clinical trials with positive enrollment, and total enrollment numbers), and finally a variable indicating if the source for that year received a CTSA award.

Figure 2 (focusing on the histograms of the “Actual” counts) shows that the distributions of the number of clinical trials changes year-by-year and for many of the years represent a zero-inflated distribution with an abundance of zero counts. Especially in the early years of the ClinicalTrials.gov collection, many sources either engage in clinical trials or simply did not register their trials at ClinicalTrials.gov, which suggests the use of a zero-inflated model. The software program Just Another Gibbs Sampler (JAGS) was used to fit independent Bayesian zero-inflated Poisson models to each of the three primary measures. The variables of source, year, and source group were incorporated into the model as regression parameters to inform the zero-inflation component and the mean parameter for the Poisson distribution. More specifically, this is the

specific Bayesian zero-inflated Poisson regression model with non-informative priors.

$$Y_i \sim \text{ZIP}(\lambda_i, \theta_i) : Y_i = 0 \text{ with probability } \theta_i \text{ otherwise } Y_i \sim \text{Poisson}(\lambda_i).$$

Bernoulli model for zero counts: Bernoulli $(1 - \theta_i)$.

$$\begin{aligned} \text{logit}(\theta_i) = & \beta_{\theta,0} + \sum_{\text{source}} \beta_{\theta,\text{source}} X_{\text{source}} \\ & + \sum_{\text{year}} \beta_{\theta,\text{year}} X_{\text{year}} + \sum_{\text{category}} \beta_{\theta,\text{group}} X_{\text{group}}. \end{aligned}$$

$$\begin{aligned} \log(\lambda_i) = & \beta_{\lambda,0} + \sum_{\text{source}} \beta_{\lambda,\text{source}} X_{\text{source}} \\ & + \sum_{\text{year}} \beta_{\lambda,\text{year}} X_{\text{year}} + \sum_{\text{category}} \beta_{\lambda,\text{group}} X_{\text{group}} \end{aligned}$$

$$\beta_{\theta,0}, \beta_{\theta,\text{source}}, \beta_{\theta,\text{year}}, \beta_{\theta,\text{group}}, \beta_{\lambda,0}, \beta_{\lambda,\text{source}}, \beta_{\lambda,\text{year}}, \beta_{\lambda,\text{group}} \sim N(0, 10,000).$$

In the above model, the matrices X_{source} , X_{year} , and X_{group} contain the respective indicator variables, so each matrix has 6,536 rows and 343, 18, and 4 columns, respectively. Altogether the model contains $2 \times (343 + 18 + 4 + 1) = 732$ parameters with flat $N(0, 10,000)$ prior distributions. The JAGS program was run with two chains and 2,000 iterations per chain with 1,000 iterations for burn-in.

Separate Bayesian models were fit for each of the three primary measures, and, for each model, predictive distributions for all 6,536 values were generated within the JAGS routine. The statistics programs R (version 3.6.0) and JAGS (version 4.3.0) with packages R2jags,⁴ ggplot2,⁵ and plotrix⁶ were used to analyze and graph the results.

RESULTS

The Bayesian models predict the number of registered clinical trials, the number of registered clinical trials with positive enrollment, and the clinical trial enrollment numbers for each institution. These predictions are informed by data available from the same institution for other years, data from other institutions in the same year, and data from other institutions in the same grouping (e.g., other CTSA institutions). In the modeling, the data values for institutions receiving CTSA funding were replaced with missing values so that effect of the CTSA funding can be assessed by comparing the actual values to the predicted values.

As previously mentioned, **Figure 2** shows how a zero-inflated Poisson distribution seems suitable for the log-transformed data. This figure also depicts the predicted values (based on the medians of the posterior distributions). The predicted distributions also capture the zero-inflated nature of the distributions. It is noted that the data for CTSA institution grouping (first row of graphs in **Figure 2**) incorporates all data from institutions that ever had CTSA funding, not just data from their funded years.

Figure 3 shows one of the main results of this analysis. In this figure, the actual and predicted log-transformed values for each of the three primary measures are graphed over time for each group. Additionally, the average percentage difference between the actual values and predicted values over time are included in the lower righthand corner of each graph. These graphs show the predicted values closely match the actual values for all groups except for the CTSA groups. In particular, based on this analysis, we conclude that CTSA funding induced approximately a 14% increase in clinical trials and a 16% increase in clinical trial enrollment. **Figure 4** isolates out just the CTSA institutions from **Figure 3** and shows the effects of the CTSA awards on just

these institutions. In particular, the figure highlights the results since 2015, in which the CTSA institutions have observed a 20% increase in clinical trials and a 42% increase in clinical trial enrollment.

Figure 5 graphs the effects of the CTSA funding as depicted by the percentage difference between actual and predicted registered clinical values against the number of years of CTSA funding. This graph shows the effect of CTSA funding on registered clinical trials is magnified with longer terms of funding.

DISCUSSION

In this report, we utilize a Bayesian zero-inflated Poisson regression model to assess the impact of the CTSA program on institutionally sponsored clinical trials. Although it is not possible to directly estimate the effect of these awards on clinical trial registration, this approach provides a reasonable approximation of the effect of the awards on the number of registered clinical trials and overall clinical trial enrollment. In particular, we find that the CTSA program, among its many diverse missions, has been successful in moving scientific innovation into clinical practice as demonstrated by an approximate 14% increase in the number of clinical trials. Furthermore, this effect increases to over 25% for institutions that have been funded for over 12 years.

This assessment utilizes data from the same institution for other years, data from other institutions in the same year, and data from other institutions in the same grouping class to predict the number of clinical trials. It is not possible to identify or control for every possible confounding effect, but the accurate predictions of the data for the non-CTSA institutions gives some credence to the prediction model used.

The CTSA program has many missions toward improving clinical and translational research, and it is expected that these efforts will eventually improve the overall quality, not just quantity, of the clinical trial research. Although this study identifies an impact of the CTSA program on registered clinical trials and clinical trial enrollment, this analysis is unable to assess the quality of the clinical trials. As the models that were fit involved many variables, we sought to include as much data as possible, which included data on trials that

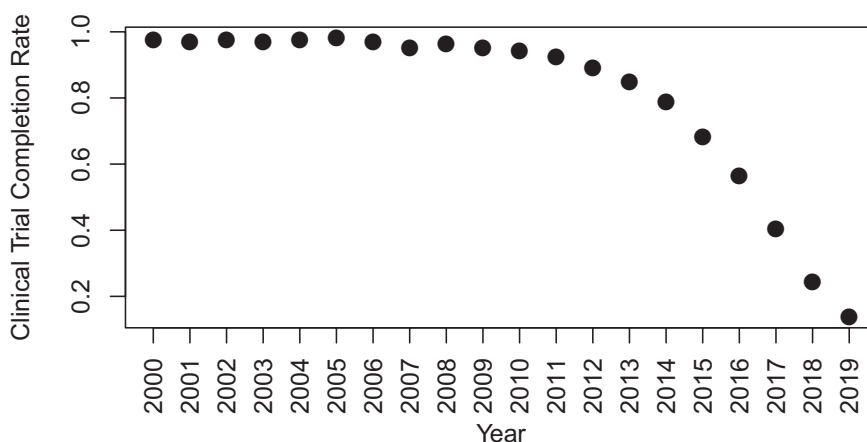


Figure 6 Graph of the completion rates of the ClinicalTrials.gov registered trials from 2000 to 2019.

were recruiting as well as completed. **Figure 6** shows the completion rates of the clinical trials per year. Although studies that were initiated prior to 2013 had a completion rate of over 89%, studies that were started between 2017 and 2019 had only 26% of the trials were listed as completed.

Supporting Information. Supplementary information accompanies this paper on the *Clinical and Translational Science* website (www.cts-journal.com).

Funding. No funding was received for this work.

Conflict of Interest. The author declared no competing interests for this work.

Author Contributions. A.B. designed the research, performed the research, analyzed the data, and wrote the manuscript.

1. Llewellyn, N., Carter, D.R., DiazGranados, D., Pelfrey, C., Rollins, L. & Nehl, E.J. Scope, influence, and interdisciplinary collaboration: the publication portfolio of the NIH

Clinical and Translational Science Awards (CTSA) Program from 2006 through 2017.

Eval. Health Prof. <https://doi.org/10.1177/0163278719839435>. [e-pub ahead of print].

2. National Center for Research Resources. Clinical and Translational Science Awards [Internet] <https://web.archive.org/web/20100527171210/http://www.ncrr.nih.gov/publications/pdf/ctsa_factsheet.pdf> (2009).
3. Liu, G., Chen, G., Sinoway, L.I. & Berg, A. Assessing the impact of the NIH CTSA program on institutionally sponsored clinical trials. *Clin. Translat. Sci.* **6**, 196–200 (2013).
4. Su, Y.-S. & Yajima, M. R2jags: Using R to run 'JAGS'. R package version 5–7. **34** (2015).
5. Wickham, H. *ggplot2: elegant graphics for data analysis*. (Springer, New York, NY, 2016).
6. Lemon, J. Plotrix: a package in the red light district of R. *R-news* **6**, 8–12 (2006).

© 2020 The Author. *Clinical and Translational Science* published by Wiley Periodicals, Inc. on behalf of the American Society of Clinical Pharmacology and Therapeutics. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.