

# A Decade On: Systematic Review of ClinicalTrials.gov Infectious Disease Trials, 2007–2017

Ian S. Jaffe,<sup>1</sup> Karen Chiswell,<sup>2</sup> and Ephraim L. Tsalik<sup>1</sup>

<sup>1</sup>Division of Infectious Diseases, Department of Medicine, Duke University School of Medicine, and <sup>2</sup>Duke Clinical Research Institute, Duke University, Durham, North Carolina

**Background.** Registration of interventional trials of Food and Drug Administration–regulated drug and biological products and devices became a legal requirement in 2007; the vast majority of these trials are registered in ClinicalTrials.gov. An analysis of ClinicalTrials.gov offers an opportunity to define the clinical research landscape; here we analyze 10 years of infectious disease (ID) clinical trial research.

**Methods.** Beginning with 166 415 interventional trials registered in ClinicalTrials.gov from 2007–2017, ID trials were selected by study conditions and interventions. Relevance to ID was confirmed through manual review, resulting in 13 707 ID trials and 152 708 non-ID trials.

**Results.** ID-related trials represented 6.9%–9.9% of all trials with no significant trend over time. ID trials tended to be more focused on treatment and prevention, with a focus on testing drugs, biologics, and vaccines. ID trials tended to be large, randomized, and nonblinded with a greater degree of international enrollment. Industry was the primary funding source for 45.2% of ID trials. Compared with the global burden of disease, human immunodeficiency virus/AIDS and hepatitis C trials were overrepresented, and lower respiratory tract infection trials were underrepresented. Hepatitis C trials fluctuated, keeping with a wave of new drug development. Influenza vaccine trials peaked during the 2009 H1N1 swine influenza outbreak.

**Conclusions.** This study presents the most comprehensive characterization of ID clinical trials over the past decade. These results help define how clinical research aligns with clinical need. Temporal trends reflect changes in disease epidemiology and the impact of scientific discovery and market forces. Periodic review of ID clinical trials can help identify gaps and serve as a mechanism to realign resources.

**Key words:** clinical trials; hepatitis C; infectious disease; policy.

The appropriate management of infectious diseases (IDs) requires an understanding of the risk and efficacy of medical drugs, biologics, and devices. In this context of evidence-based medicine, clinical trials contribute knowledge to replace or confirm prior dogma. As discussed elsewhere, ID specialists often contend with a lack of clinical trials (and particularly randomized controlled trials) to inform clinical decision making [1, 2].

ClinicalTrials.gov is a Food and Drug Administration (FDA)–managed database of >265 000 clinical trials from 203 nations. Registration with ClinicalTrials.gov has been a legal requirement for most interventional trials of FDA-regulated drug and biological products and devices conducted in the United States since September 2007, in addition to being a requirement for

publication in many peer-reviewed medical journals [3, 4]. In 2010, an initiative of the Clinical Trials Transformation Institute began to describe the trials represented in the ClinicalTrials.gov database from 2007 to 2010 [5]. ID trials in the database were previously described as part of this 3-year “snapshot.” [1] Trials focusing on human immunodeficiency virus (HIV) and hepatitis C, respiratory tract infections, and pediatric antibiotic and antifungal trials were subsequently reported [6–8].

In this cross-sectional study, we aimed to characterize the scope and nature of ID clinical trials in ClinicalTrials.gov for a 10-year period through a systematic analysis of registered trials. As with the 2007–2010 snapshot, we examined trial methods and funding source, with a new emphasis on analyzing trends over time. We also evaluated the alignment between current clinical research priorities and the disease burden of IDs in the United States and abroad. In addition, we provided a more detailed analysis of hepatitis C trials given the significant changes in hepatitis C management over the past decade.

## METHODS

We performed a systematic analysis of characteristics of ID trials registered with ClinicalTrials.gov from 1 October 2007 to 30 September 2017. We chose ClinicalTrials.gov for our

Received 22 January 2019; editorial decision 4 April 2019; accepted 11 April 2019.

Correspondence: Ephraim L. Tsalik, Department of Medicine, Duke University Medical Center, 157 Hanes House, DUMC 102359, Durham, NC 27710 (e.t@duke.edu).

### Open Forum Infectious Diseases®

© The Author(s) 2019. Published by Oxford University Press on behalf of Infectious Diseases Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com  
DOI: 10.1093/ofid/ofz189

analysis because it was used for the 2007–2010 snapshot and, as described elsewhere, it is among the largest trial repositories, has comprehensive data fields, and has established methods for data download [1]. Although ClinicalTrials.gov is not the most comprehensive database of clinical trials (compared with the International Clinical Trials Registry Platform), it is an accurate reflection of clinical trials being conducted in IDs and is amenable to aggregate analysis [1]. Moreover, a comprehensive review of ID trials has not been completed since the initial snapshot, although reviews have been published of respiratory tract infection trials, HIV and hepatitis C trials, and pediatric antibacterial and antifungal trials [6–8]. The 10 years of clinical trials examined in this study include trials submitted to comply with requirements for publication in peer-reviewed medical journals and trials registered under the FDA Amendments Act of 2007 and other regulatory requirements [9].

### Creation of the ID Study Data Set

The ID study data set was created as described elsewhere [1]. We used the 16 October 2017 version of the database for Aggregate Analysis of ClinicalTrials.gov to identify clinical trials registered with ClinicalTrials.gov from 1 October 2007 through 30 September 2017. This cohort represents the entirety of clinical trials registered after US legal requirements for registering certain interventional trials available at the time of this study [9]. Aggregate Analysis of ClinicalTrials.gov is a relational database (PostgreSQL) developed and maintained by the Clinical Trials Transformation Initiative. It contains all information about studies registered in ClinicalTrials.gov since its inception in February 2000 and is updated daily with content downloaded from ClinicalTrials.gov. The database is publicly available in the cloud, with access information and documentation provided at the Clinical Trials Transformation Initiative website [10]. We then focused on interventional trials by filtering the data set using the registry’s “study type” field, which identifies studies as interventional, observational, expanded access, or not applicable.

To identify ID trials, we focused on the condition and intervention characteristics, which were defined by submitted data and linked Medical Subject Heading (MeSH) terms generated by a National Library of Medicine (NLM) algorithm based on the 2017 MeSH thesaurus [11]. Of 10 466 MeSH terms manually reviewed, 991 (9.5%) were related to IDs. Because some conditions could not be linked to MeSH terms, free-text condition terms appearing in  $\geq 5$  interventional trials were also manually reviewed for relevance. Of 5140 possible free-text condition terms reviewed, 345 were relevant to IDs (6.7%) (Supplementary Table 1).

ID trials were also identified using the submitted intervention term. Intervention terms linked to MeSH terms generated by the NLM algorithm that appeared in  $\geq 4$  clinical trials were reviewed for relevance. Of the 2101 intervention MeSH terms reviewed, 309 were identified as relevant to IDs (14.7%). An

initial data set of 19 794 trials was generated by identifying trials with  $\geq 1$  relevant term in the NLM-generated MeSH condition field, the submitted free-text condition field, or the submitted intervention name field. This process was previously developed and validated by comparison with classifications based on manual review, for studies of cardiovascular diseases, cancer, and mental health [12]. Trials were then manually reviewed by one of us (I. S. J.) to exclude non-ID studies. A total of 13 707 ID studies were identified, which defines the study data set used for this analysis.

### Subcategorization of the ID Study Data Set

After defining the ID trials data set, we subcategorized trials by IDs based on study title and description. Subcategories were defined based on World Health Organization (WHO) cause-of-death groupings, excluding “maternal conditions” and “perinatal conditions” [1]. Along with the 18 WHO categories, 42 additional categories were defined (for a total of 60), such that each trial was assigned to  $\geq 1$  subcategory. Trials that fit equally well into multiple subcategories were assigned to multiple categories. Trials were also manually categorized as vaccine trials or nonvaccine trials. The percentage of all ID-related deaths and ID-related disability-adjusted life-years (DALYs) attributable to selected conditions was calculated from the 2000–2015 WHO Global Burden of Disease [13, 14].

### Analytical Methods

We used R software, version 3.4 (Foundation for Statistical Computing) to calculate frequencies and percentages for categorical trial characteristics and median and interquartile ranges (IQRs) for continuous characteristics. Disease prevalence and disease-specific DALYs were derived from the 2015 WHO global health estimates summary tables [13]. Actual enrollment or anticipated enrollment was reported and summary statistics were calculated by pooling across active and completed trials. Probable funding source (because ClinicalTrials.gov does not require funding source be reported) was attributed based on the lead sponsor and collaborator fields and reported as “industry,” “NIH” [National Institutes of Health], “U.S. federal (excluding NIH),” or “Other.” A trial was considered “industry-funded” if the lead sponsor was from industry, or if the NIH was neither a lead sponsor nor collaborator and  $\geq 1$  collaborator was from industry. An “NIH-funded” study required the NIH to be either a lead sponsor or a collaborator, and no industry as lead sponsor. “Other” was used to describe studies for which the lead sponsor and collaborator fields were nonmissing and did not meet criteria for either industry or NIH funding. Countries were grouped into 11 global regions to allow analysis of geographical regions.

## RESULTS

The initial data set downloaded on 16 October 2017 included 256 544 clinical trials registered with ClinicalTrials.gov.

A total of 166 415 interventional trials were registered from 1 October 2007, after enactment of mandatory registration on 27 September 2007, through 30 September 2017, providing a 10-year period for studying trends in registered trials. Of these, 13 707 trials (8.5%) were defined as the ID trial data set.

A summary of ID trials, non-ID trials, and selected subcategories (HIV/AIDS, hepatitis C, malaria, skin and soft-tissue infection, and lower respiratory tract infection [LRTI]) is presented in Table 1 [15]. These categories were chosen for further analysis based on their prevalence as well as global and domestic significance. The primary purpose was “treatment” in the majority of both ID (52.7%) and non-ID trials (69.6%). A higher proportion of ID trials than non-ID trials focused on prevention (32.5% vs 9.1%, respectively). In terms of intervention type, ID trials also tended to be more drug focused than non-ID trials (54.4% vs 50.3%, respectively) and more biologic/vaccine focused (23.5% vs 5.0%). Of 3215 ID trials with a biologic or vaccine intervention, 85.3% evaluated a vaccine. ID trials were less likely than non-ID trials to concern interventions involving procedures (4.2% vs 10.7%, respectively) or devices (5.9% vs 14.2%).

ID trials also tended to be larger than non-ID trials, with a median (interquartile range) enrollment of 102 (37–321.5) versus 60 (27–145) subjects for non-ID studies. The largest studies were observed for *Enterovirus* trials (mean enrollment, 6855 subjects; median, 780), trachoma (mean, 5035; median 1139), intestinal nematode infection (mean, 4323; median, 172), and malaria (mean, 3652; median, 141). These results were due primarily to the inclusion of multiple large (>10 000-subject) studies, including 5 *Enterovirus*, 1 trachoma, 2 intestinal nematode infection, and 28 malaria trials. Five trials included ≥100 000 subjects, 1 for intestinal nematode infection trial and 4 for malaria. However, only the *Enterovirus* and trachoma categories had median enrollments of >500 subjects.

A greater percentage of ID trials focused specifically on children (≤18 years of age (13.3%), compared with non-ID trials (6.3%). Moreover, ID trials were more likely than non-ID trials to exclude elderly patients (>65 years of age) (32.8% vs 2.9%, respectively). After omission of trials restricted to pediatric subjects, ID trials were still more likely than non-ID trials to exclude elderly patients (21.3% vs 17.8%, respectively). Omitting trials restricted to pediatrics, we observed that certain trial categories excluded elderly patients at higher rates: pharmacology (54.5%), HIV/AIDS (31.4%), influenza vaccine (29.8%), malaria (28.6%), and tuberculosis (21.7%) trials were more likely to exclude the elderly. In line with non-ID trials, the majority of ID trials were randomized (85.9%) and did not use masking protocols (57.1%). ID trials were less likely than non-ID trials to occur in North America (37.8% vs 44.9%) or Europe (24.5% vs 29.2%).

Of the 13 707 ID trials, 11 881 (86.7%) were assigned to 1 clinical subcategory, 1754 (12.8%) were assigned to 2, and 72 (0.5%)

were assigned to ≥3. The distribution of trials across the 60 ID subcategories is presented in Figure 1. The 4 most common trial categories (HIV/AIDS, hepatitis C, influenza vaccine, and LRTI) accounted for 35.6% of all ID trials, in terms of trials assigned solely to those categories (4875 trials; Figure 1). Figure 1 also shows the total number of subjects enrolled or expected to be enrolled in each ID subcategory. HIV/AIDS had the highest numbers of trials and enrolled subjects. Despite having fewer trials, malaria and tuberculosis had the second and fourth highest numbers of enrolled subjects, owing to relatively large enrollment sizes.

Based on their frequency in the database and their global impact, 10 subcategories were chosen for more detailed characterization. The representation of these selected subcategories within the ClinicalTrials.gov ID data set was compared with global and United States ID-related deaths and DALYs (as defined by the WHO Global Burden of Disease) and is shown in Figure 2. Hepatitis C trials accounted for 8.7% of ID trials registered from 2007–2017. This was similar to the hepatitis C virus–associated morbidity rate in the United States (6.3%) but higher than the global hepatitis C virus–associated morbidity rate (0.7%). Similarly, trials of HIV/AIDS, hepatitis B, and sexually transmitted disease excluding HIV were overrepresented compared with both global and US burden of disease, whereas LRTI trials were underrepresented. Notably, diarrheal disease and tuberculosis trials were underrepresented compared with the global burden of disease (4.1% vs 16.7%, and 3.5% vs 11.0%, respectively).

Figure 3 shows the number of trials registered per study year (1 October of the year listed through 30 September of the following year) for 7 subcategories: hepatitis B, influenza vaccine, HIV/AIDS, malaria, LRTI, hemorrhagic viruses, and Zika virus. These subcategories were selected either because of interesting trends, as in the case of HIV/AIDS, influenza vaccine, hemorrhagic viruses, and Zika virus, or as representative examples for other disease categories, as in the case of hepatitis B, malaria, and LRTI. The number of trials registered per year remained steady for hepatitis B and malaria. There was a 64.0% increase in influenza vaccine studies in 2008 compared with 2007. This increase was sustained for 2 years and then declined to prior levels, where it remained stable. A 4.7-fold increase in hemorrhagic virus trials was observed in study year 2014. Ebola trials made up 32 (68.1%) of the 47 hemorrhagic virus trials registered from October 2014 through September 2015. No Zika virus trials were registered before 2015; 1 was registered in study year 2015, and 8 in 2016. The number of HIV/AIDS trials registered in any given year fluctuated from 250 to 342 but overall remained similar over the 10-year study window. The number of LRTI trials trended upward toward the end of the study period, with an increase of 97.0% in trials registered in the last 2 study years compared with the first 2.

Hepatitis C trials were selected for additional scrutiny owing to the considerable advancements in hepatitis C treatment over the 10-year study period [16]. Hepatitis C trials tended to be

**Table 1. Characteristics of Infectious Disease (ID) Studies, Non-ID Studies, and Selected ID Subcategories**

Parameter <sup>a</sup>	Subjects by Study Focus, No. (%) <sup>b</sup>						
	Non-ID (n = 152 708)	All ID (n = 13 707)	HIV/AIDS (n = 3037)	Hepatitis C (n = 1189)	Malaria (n = 562)	SSTI (n = 874)	LRTI (n = 902)
Primary purpose <sup>c</sup>	(n = 145 135)	(n = 13 159)	(n = 2891)	(n = 1125)	(n = 545)	(n = 849)	(n = 871)
Treatment	100 956 (69.6)	6938 (52.7)	1484 (51.3)	1001 (89.0)	259 (47.5)	711 (83.7)	519 (59.6)
Prevention	13 159 (9.1)	4279 (32.5)	823 (28.5)	35 (3.1)	199 (36.5)	73 (8.6)	204 (23.4)
Other purpose	31 020 (21.4)	1942 (14.8)	584 (20.2)	89 (7.9)	87 (16.0)	65 (7.7)	148 (17.0)
Intervention <sup>d</sup>							
Drug	76 774 (50.3)	7456 (54.4)	1640 (54.0)	1059 (89.1)	374 (66.5)	567 (64.9)	566 (62.7)
Procedure	16 315 (10.7)	571 (4.2)	60 (2.0)	15 (1.3)	11 (2.0)	70 (8.0)	38 (4.2)
Biological/vaccine	7566 (5.0)	3215 (23.5)	403 (13.3)	89 (7.5)	137 (24.4)	41 (4.7)	117 (13.0)
Device	21 615 (14.2)	806 (5.9)	91 (3.0)	16 (1.3)	20 (3.6)	177 (20.3)	76 (8.4)
Other	42 354 (27.7)	2163 (15.8)	897 (29.5)	84 (7.1)	59 (10.5)	78 (8.9)	148 (16.4)
Vaccine <sup>e</sup>	NA	2743 (20.0)	332 (10.9)	17 (1.4)	87 (15.5)	5 (0.6)	58 (6.4)
Lead sponsor							
Industry	43 581 (28.5)	5074 (37.0)	751 (24.7)	738 (62.1)	92 (16.4)	473 (54.1)	327 (36.3)
NIH	1761 (1.2)	469 (3.4)	165 (5.4)	20 (1.7)	30 (5.3)	9 (1.0)	35 (3.9)
US federal	1686 (1.1)	193 (1.4)	29 (1.0)	15 (1.3)	34 (6.0)	9 (1.0)	2 (0.2)
Other	105 680 (69.2)	7971 (58.2)	2092 (68.9)	416 (35.0)	406 (72.2)	383 (43.8)	538 (59.6)
Funder							
Industry	55 464 (36.3)	6198 (45.2)	1120 (36.9)	856 (72.0)	133 (23.7)	516 (59.0)	388 (43.0)
NIH	9716 (6.24)	1221 (8.9)	663 (21.8)	52 (4.4)	52 (9.3)	17 (1.9)	63 (7.0)
Other	87 528 (57.3)	6288 (45.9)	1254 (41.3)	281 (23.6)	377 (67.1)	341 (39.0)	451 (50.0)
Enrollment	(n = 152 010)	(n = 13 643)	(n = 3026)	(n = 1183)	(n = 557)	(n = 866)	(n = 898)
Subjects, median (IQR), No.	60 (27–145)	102 (37–321.5)	88 (30–333)	61 (27–163)	141 (38–600)	80 (30–212)	100 (40–300)
Sex/age eligibility							
Female	15 115 (9.9)	976 (7.1)	441 (14.5)	12 (1.0)	44 (7.8)	85 (9.7)	16 (1.8)
Male	8738 (5.7)	546 (4.0)	283 (9.3)	23 (1.9)	27 (4.8)	6 (0.7)	15 (1.7)
Both	128 855 (84.4)	12 185 (88.9)	2313 (76.2)	1154 (97.1)	491 (87.4)	783 (89.6)	871 (96.6)
Children only <sup>f</sup>	9573 (6.3)	1823 (13.3)	168 (5.5)	13 (1.1)	142 (25.3)	64 (7.3)	170 (18.8)
Excluding elderly subjects <sup>g</sup>	35 028 (22.9)	4354 (32.8)	1122 (36.9)	182 (15.3)	303 (53.9)	163 (18.6)	205 (22.7)
Allocation	(n = 115 919)	(n = 11 369)	(n = 2436)	(n = 888)	(n = 491)	(n = 726)	(n = 773)
Randomized	100 271 (86.5)	9766 (85.9)	2065 (84.8)	667 (75.1)	403 (82.1)	675 (93.0)	692 (89.5)
Nonrandomized	15 648 (13.5)	1603 (14.1)	371 (15.2)	221 (24.9)	88 (17.9)	51 (7.0)	81 (10.5)
Masking	(n = 151 476)	(n = 13 632)	(n = 3019)	(n = 1184)	(n = 559)	(n = 866)	(n = 898)
Open	84 851 (56.0)	7787 (57.1)	2071 (68.6)	866 (73.1)	379 (67.8)	382 (44.1)	414 (46.1)
Single	21 270 (14.0)	1300 (9.5)	250 (8.3)	28 (2.4)	47 (8.4)	128 (14.8)	81 (9.0)
Double	45 355 (29.9)	4545 (33.3)	698 (23.1)	290 (24.5)	133 (23.8)	356 (41.1)	403 (44.9)
No. of arms	(n = 151 679)	(n = 13 636)	(n = 3013)	(n = 1174)	(n = 561)	(n = 870)	(n = 893)
1	41 994 (27.7)	2750 (20.1)	707 (23.5)	341 (29.0)	83 (14.8)	158 (18.2)	150 (16.8)
2	80 668 (53.2)	7218 (53.0)	1591 (52.8)	424 (36.1)	238 (42.4)	515 (59.2)	553 (61.9)
3+	29 017 (19.1)	3668 (26.9)	715 (23.7)	409 (34.8)	240 (42.8)	197 (22.6)	190 (21.3)
Phase							
Early phase 1	1838 (1.2)	100 (0.7)	29 (1.0)	7 (0.6)	3 (0.5)	8 (0.9)	12 (1.3)
Phase 1	19 967 (13.1)	2280 (16.6)	599 (19.7)	253 (21.3)	134 (23.8)	77 (8.8)	136 (15.1)
Phase 1/phase 2	6690 (4.4)	517 (3.8)	129 (4.2)	40 (3.4)	43 (7.7)	31 (3.5)	25 (2.8)
Phase 2	24 805 (16.2)	2252 (16.6)	443 (14.6)	300 (25.2)	62 (11.0)	158 (18.1)	141 (15.6)
Phase 2/phase 3	3158 (2.1)	355 (2.6)	74 (2.4)	26 (2.2)	19 (3.4)	25 (2.9)	24 (2.7)
Phase 3	16 438 (10.8)	2298 (17.5)	400 (13.2)	272 (22.9)	83 (14.8)	175 (20.0)	156 (17.3)
Phase 4	16 125 (10.6)	2400 (17.5)	413 (13.6)	186 (15.6)	110 (19.6)	146 (16.7)	162 (18.0)
NA	63 687 (41.7)	3405 (24.8)	950 (31.3)	105 (8.8)	108 (19.2)	254 (29.1)	246 (27.3)
Overall status <sup>h</sup>							
Not yet recruiting	8175 (5.4)	621 (4.5)	124 (4.1)	32 (2.7)	12 (2.1)	51 (5.8)	49 (5.4)
Recruiting	33 598 (22.0)	2204 (16.1)	489 (16.1)	114 (9.6)	66 (11.7)	179 (20.5)	183 (20.3)
Active, not recruiting	11 206 (7.3)	902 (6.6)	302 (9.9)	57 (4.8)	23 (4.1)	32 (3.7)	51 (5.7)
Completed	73 116 (47.9)	7872 (57.4)	1724 (56.8)	801 (67.4)	375 (66.7)	468 (53.5)	454 (50.3)
Terminated	4381 (2.9)	389 (2.8)	71 (2.3)	46 (3.9)	16 (2.8)	25 (2.9)	39 (4.3)
Unknown	22 182 (14.5)	1719 (12.5)	327 (10.8)	139 (11.7)	70 (12.5)	119 (13.6)	126 (14.0)



**Table 1. Continued**

Parameter <sup>a</sup>	Subjects by Study Focus, No. (%) <sup>b</sup>						
	Non-ID (n = 152 708)	All ID (n = 13 707)	HIV/AIDS (n = 3037)	Hepatitis C (n = 1189)	Malaria (n = 562)	SSTI (n = 874)	LRTI (n = 902)
Regional distribution <sup>c</sup>							
Africa	3150 (2.1)	1333 (9.7)	534 (17.6)	45 (3.8)	286 (50.9)	32 (3.7)	56 (6.2)
Central America	1076 (0.7)	427 (3.1)	136 (4.5)	127 (10.7)	0 (0.0)	38 (4.3)	22 (2.4)
North America	68 615 (44.9)	5187 (37.8)	1430 (47.1)	518 (43.6)	59 (10.5)	489 (56.0)	324 (35.9)
South America	5702 (3.7)	637 (4.6)	151 (5.0)	51 (4.3)	17 (3.0)	32 (3.7)	70 (7.8)
East Asia	18 029 (11.8)	1479 (10.8)	145 (4.8)	148 (12.4)	4 (0.7)	36 (4.1)	114 (12.6)
North Asia	3116 (2.0)	296 (2.2)	71 (2.3)	46 (3.9)	0 (0.0)	17 (1.9)	46 (5.1)
South Asia	2388 (1.6)	403 (2.9)	45 (1.5)	20 (1.7)	16 (2.8)	17 (1.9)	37 (4.1)
Southeast Asia	3015 (2.0)	690 (5.0)	145 (4.8)	23 (1.9)	79 (14.1)	17 (1.9)	52 (5.8)
Europe	44 516 (29.2)	3360 (24.5)	625 (20.6)	347 (29.2)	72 (12.8)	173 (19.8)	322 (35.7)
Middle East	6262 (4.1)	386 (2.8)	12 (0.4)	48 (4.0)	1 (0.2)	31 (3.5)	47 (5.2)
Pacifica	3817 (2.5)	476 (3.5)	72 (2.4)	129 (10.8)	21 (3.7)	12 (1.4)	57 (6.3)
Unknown	16 426 (10.8)	1586 (11.6)	339 (11.2)	213 (17.9)	31 (5.5)	96 (11.0)	103 (11.4)

Abbreviations: HIV, human immunodeficiency virus; ID, infectious disease; IQR, interquartile range; LRTI, lower respiratory tract infection; NA, not applicable; NIH, US National Institutes of Health; SSTI, skin and soft-tissue infection.

<sup>a</sup>The denominator for each variable is the number of trials reporting such data.

<sup>b</sup>Data represent no. (%) of subjects unless otherwise specified.

<sup>c</sup>“Other purpose” includes diagnostic, supportive care, screening, health services research, and basic science.

<sup>d</sup>For intervention types, the numerator is the number of trials with ≥1 intervention of this type. A study may have multiple intervention types; hence, cumulative percentages may not equal 100%. “Other” interventions include radiation, dietary supplement, and genetic.

<sup>e</sup>Trials were manually determined to have involved a vaccine. Non-ID studies by definition could not include any vaccine trials.

<sup>f</sup>Children were defined as subjects ≤18 years of age.

<sup>g</sup>Elderly subjects were defined as those >65 years of age.

<sup>h</sup>The recruiting category includes trials recruiting by invitation; the terminated category includes trials that have been terminated, suspended, or withdrawn.

<sup>i</sup>The numerator for the regional distribution variable is the number of trials with ≥1 trial site in that region. Trials can be in multiple regions; hence, the cumulative percentages may exceed 100%. Country regions are available at the ClinicalTrials Transformation Initiative website [15].

more focused on treatment (89.0%) and specifically drug intervention (89.1%) than the rest of the ID trials (54.5%). They were also more likely to be funded by industry (72.0% vs 40.5%). Hepatitis C trials registered by year were pooled by phase as follows: early phase 1 trials were considered to be phase 1, phase 1/phase 2 trials were considered phase 2, and phase 2/phase 3 trials were considered phase 3. The distribution of the phases of hepatitis C trials is presented in Figure 4, along with key dates for these drugs, including date of first registration in ClinicalTrials.gov and date of FDA approval [17–20]. As these new hepatitis C drugs entered into clinical testing, there is a phase shift evident in the trial phase, which begins with a predominance of phase 1, then advancing through the subsequent phases over time. Funding sources for hepatitis C trials per year are shown in Figure 4B. In general, industry funding plateaued in study years 2011 and 2012, with industry being the primary funding source for >80% of trials and dropping to 52.9% of studies by 2011. NIH funding fluctuated, but the NIH was the main funder for <10% of hepatitis C trials for all study years.

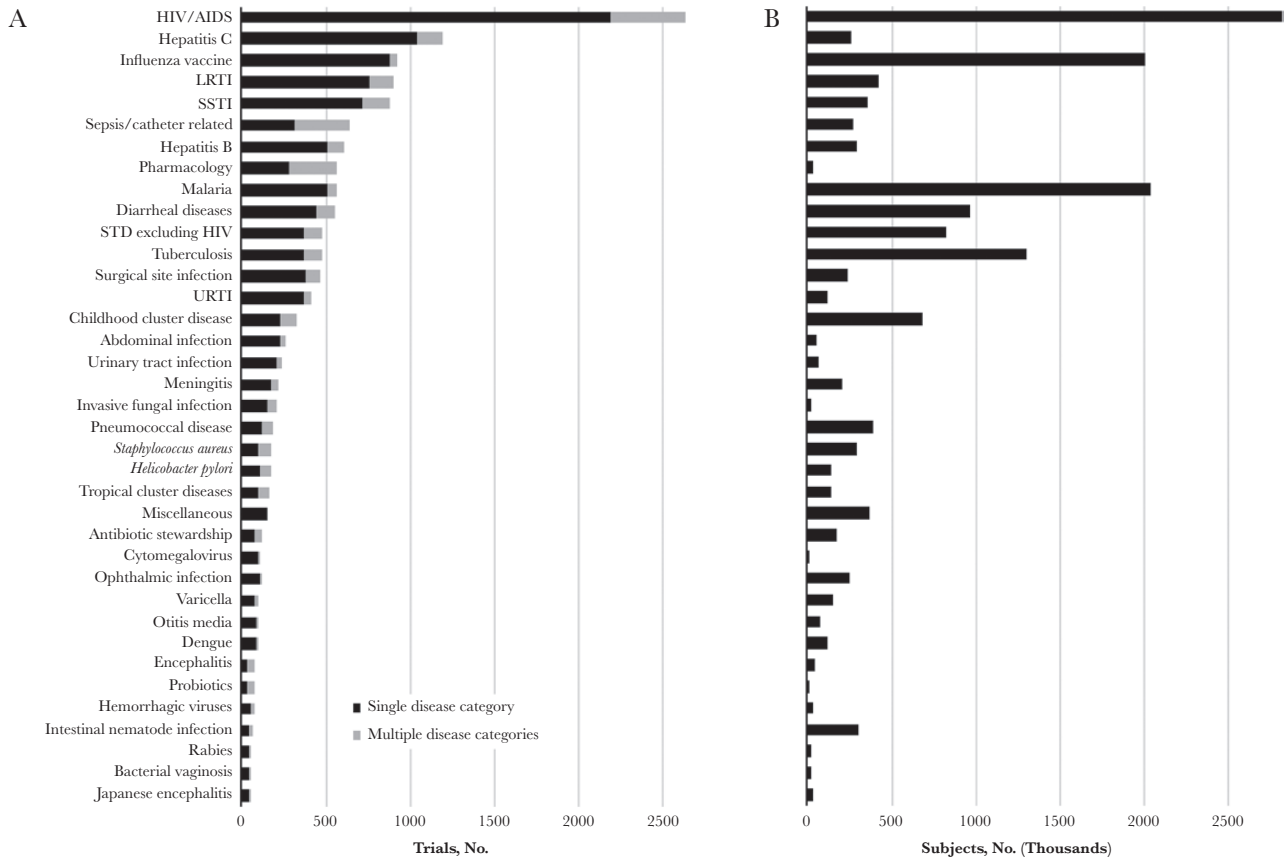
## DISCUSSION

The ClinicalTrials.gov registry serves as one of the largest repositories of clinical trials. Although originally intended to be a portal that links clinical trials to the general public, it also

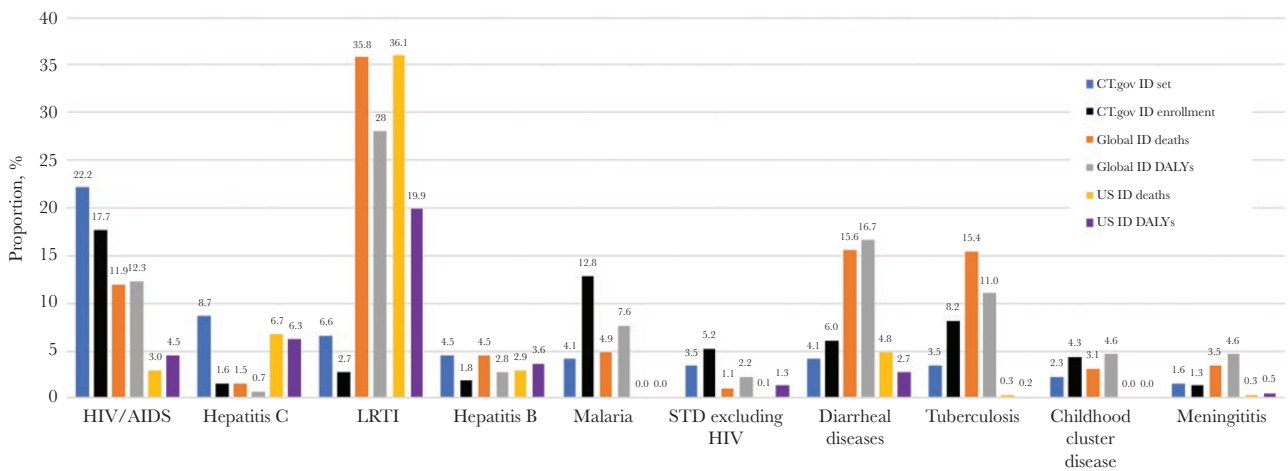
serves as a useful tool to understand the clinical trial landscape. Several analyses of the registry have been published, focusing on overall trial characteristics, disease-specific characteristics, and compliance with results reporting, among other areas [5, 21, 22]. We previously used this database to characterize the landscape of ID trials, although that analysis was limited to the first 3 years (2007–2010) after registration of interventional trials became legally mandatory. Now that a decade of ClinicalTrials.gov data after FDA Amendments Act registration mandates are available, there is an opportunity to reanalyze the ID research landscape.

Compared with our prior analysis of 2007–2010 trial data, many characteristics remained relatively unchanged [1]. This is largely because of qualities inherent to IDs. For example, there was a strong emphasis on prevention-focused trials. Similarly, the importance of vaccine-based interventions persisted. ID trials also restricted trials to children at a much higher rate than non-ID trials, potentially owing to the focus on prevention and vaccination within the field. This may be due to the greater burden of certain IDs in children. This predilection of IDs for the young partially, but not entirely, explains the higher rate of exclusion for elderly subjects. The exclusion of elderly subjects in the HIV/AIDS, influenza vaccine, tuberculosis, and malaria categories may limit the generalizability of some trials within these categories.

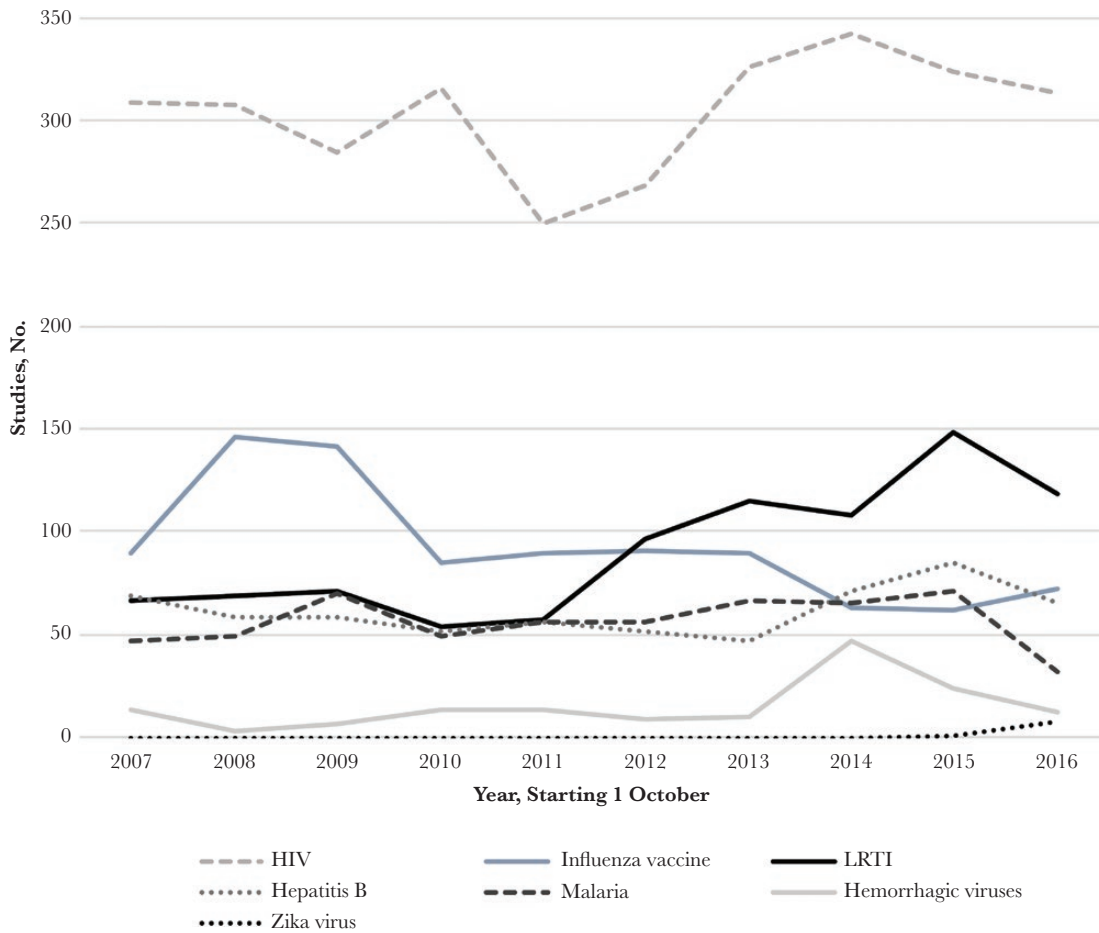
Despite registration requirements in ClinicalTrials.gov focusing on US-based and funded studies, ID trials tended to be



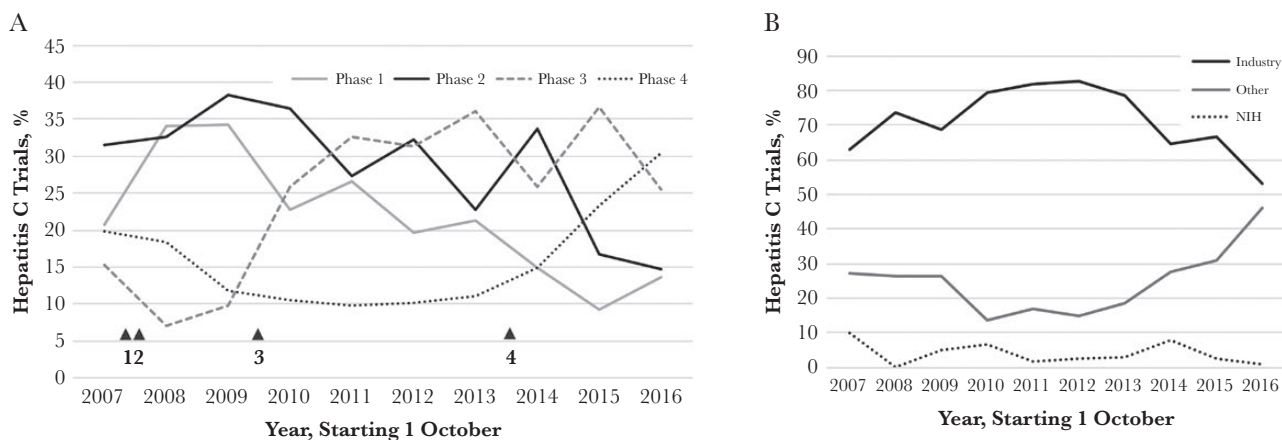
**Figure 1.** Number of infectious disease (ID) studies and enrollment by subcategory. A. Number of trials in 36 of the 60 ID subcategories in the ID trials data set. There were 24 subcategories with <32 trials over the 10-year study period that are not represented here; they include *Haemophilus*, infective endocarditis, osteomyelitis, virus as therapy, hepatitis A, anthrax, Epstein-Barr virus, nontuberculous mycobacteria, otitis externa, Lyme disease, *Enterovirus*, smallpox, hand foot mouth disease, BK virus, yellow fever, leprosy, human T-lymphotropic virus, Zika virus, *Toxoplasma*, trachoma, Q fever, Bell palsy, and tularemia. B. The total number of subjects enrolled or expected to be enrolled in studies in the same 36 subcategories. Abbreviations: HIV, human immunodeficiency virus; LRTI, lower respiratory tract infection; SSTI, skin and soft-tissue infection; STD, sexually transmitted disease; URTI, upper respiratory tract infection;



**Figure 2.** Representation of infectious disease (ID) trials and burden of disease for selected disease categories. The percentage of selected ID subcategories registered in ClinicalTrials.gov and the total enrollments for these subcategories are compared with the burden of disease globally and in the United States. The percentages of deaths and disability-adjusted life years (DALYs) are defined relative to total mortality and DALYs lost owing to IDs, excluding maternal and perinatal infections. The percent of enrollment is defined relative to total enrollment across all ID trials in the ID data set. Abbreviations: HIV, human immunodeficiency virus; LRTI, lower respiratory tract infection; STD, sexually transmitted disease.



**Figure 3.** Number of trials registered per year of selected infectious disease (ID) subcategories. Based on when interventional trial registration became legally required, years are defined as 1 October through 30 September and are based on the date of registration in ClinicalTrials.gov. Abbreviations: HIV, human immunodeficiency virus; LRTI, lower respiratory tract infection.



**Figure 4.** Hepatitis C trial funding sources and phases. *A*, Percentage of hepatitis C clinical trials registered per study year, grouped by phase. The introduction of selected hepatitis C drugs for clinical trial and Food and Drug Administration (FDA) approval of these interventions is indicated by numbers, as follows: (1) first daclatasvir trial registered on ClinicalTrials.gov (first NS5A inhibitor); (2) first simeprevir trial registered; (3) first sofosbuvir trial registered; (4) FDA approval of simeprevir, sofosbuvir, and Harvoni (sofosbuvir plus ledipasvir). *B*, Primary funding sources for hepatitis C clinical trials registered by year. The funding source was not provided by study sponsors and was instead determined algorithmically, as defined in Methods. Abbreviation: NIH, National Institutes of Health.

more global, which highlights the global effects of IDs. ID trials also tended to be larger than non-ID trials. Certain disease categories, such as malaria, had much larger trials than other ID categories. Consequently, the number of trials in a category is not always a good proxy for the extent of work being done on a particular disease. Although some disease categories did have a higher median number of enrolled subjects per trial, total enrollment numbers for some categories were skewed by the presence of large trials ( $\geq 10\,000$ ) patients, which tended to be long-term postmarketing surveillance trials.

Our analysis also found that, in the case of hepatitis C, drug and therapy development seems to follow anticipated patterns and funding sources. Hepatitis C presented a unique opportunity to observe the evolution of clinical trials research as new treatments became available. Predictably, trials progressed from early- to late-phase designs as new directly acting antiviral drugs moved through development and validation. Industry funding, the primary source of funding for these and most ID studies, tracked closely with the number of late-phase studies, indicating the large extent to which industry funding was the driving force behind hepatitis C research during this period. The new drug classes took 4–6 years to progress through the clinical research pipeline, with the first 2 years focused on early-phase trials. It is worth noting that hepatitis C trials were found to be overrepresented in the ID trial data set compared with global morbidity and mortality rates. Considering that the parallel development of multiple drug interventions was occurring at essentially the same time, it is possible some of the overrepresentation of hepatitis C could have been averted through greater collaboration of industry, funding sources, and researchers. Platform trials, such as the Investigation of Serial Studies to Predict Your Therapeutic Response With Imaging And molecular Analysis 2 (I-SPY 2) model for breast cancer, could have been an ideal opportunity to streamline evaluation of these various new hepatitis C therapies [23].

Although hepatitis C trials exhibited clear patterns, many other disease classes did not, including HIV/AIDS and malaria trials. Meanwhile, the number of LRTI trials increased toward the end of the study period, as did the number of trials focusing on sexually transmitted diseases excluding HIV, skin and soft-tissue infections, sepsis/catheter-related infections, hepatitis B, and diarrheal diseases (data not shown). Influenza vaccine trials exhibited a distinct spike in the number of clinical trials in study years 2008 and 2009, corresponding to the H1N1 swine influenza epidemic worldwide [24]. Hemorrhagic virus trials had a distinct spike in 2014–2015, corresponding to the 2014–2016 Ebola hemorrhagic fever epidemic in Western Africa [25]. Interestingly, the first Ebola case was reported in March 2014, just 6 months before the registration of several Ebola clinical trials. This is a relatively short time from the recognition of a need for clinical research, securing funding for that research, and initiation of the clinical trial [25]. A similar

trend was observed with Zika trials, which also were registered only in 2015 and 2016, corresponding to the onset of that outbreak [26]. Despite extensive research on Middle East respiratory syndrome coronavirus during 2007–2017, there were very few ( $n = 3$ ) interventional trials registered in ClinicalTrials.gov pertaining to these infections. This is probably because the registry and our analysis focuses on interventional trials, whereas many studies of Middle East respiratory syndrome coronavirus were observational and were therefore not required to register.

We found that the frequency of particular ID trial subcategories continued to poorly correlate with their global or US health impact. For example, hepatitis C trials were overrepresented compared with global and US mortality and disability rates. This should not imply that hepatitis C treatment is not clinically important. Rather, it highlights the discrepancy between the market forces that drive drug development and where clinical need may actually be greatest. Along these lines, LRTI trials continued to be underrepresented, constituting only 6.6% of the ID studies registered and 35.8% of infection-related global deaths. Malaria, diarrheal diseases, tuberculosis, childhood cluster diseases, and meningitis were also underrepresented compared with global ID morbidity and mortality rates, although these diseases were generally well represented or overrepresented compared with their very low prevalences in the United States.

The information presented herein may help identify how resources are being invested across the ID spectrum. Trial sponsors are generally focused on their respective studies and do not examine the entire portfolio of interventional ID research. A multiple-stakeholder approach to funding that incorporates the perspectives of industry, funding agencies, and policy makers may be better able to direct how resources are allocated and research areas prioritized such that public health needs are better matched to the market forces that drive much of the clinical trial enterprise.

The ID trials data set used for this analysis has several limitations. First, ClinicalTrials.gov was originally designed as a public repository for research trials and was not intended to support aggregate analysis for research purposes. As a result, trials may not be annotated by their sponsors in a consistent or complete manner. Second, the FDA requirement for registration applies to interventional trials of FDA-regulated drug and biologic products and devices. Although “intervention” encompasses a wide range of topics, a substantial amount of research, including clinical research, is noninterventional and therefore not captured in this data set [3, 27]. In addition, less common intervention types (eg, dietary and behavioral intervention) may not be captured. Third, ClinicalTrials.gov does not collect information about trial funding. We have attempted to present funding information using previously published algorithms. However, these algorithms make certain assumptions based on study sponsor, which limits some of the conclusions that can be



made about funding. Fourth, ClinicalTrials.gov does not collect information regarding drug and therapeutic approvals, limiting our ability to draw conclusions on intervention development pipelines.

This examination of 10 years of ClinicalTrials.gov data revealed that ID trials are well represented in the overall clinical research landscape. They include a large global footprint, but the distribution of studies is not consistent with the US or global burden of disease. The registry also reflected real-world changes in drug discovery, research, and approval as seen for hepatitis C. Although these observations are necessarily retrospective, understanding the history of ID trials can also inform their future and provide an opportunity to examine the mechanisms through which research is prioritized.

### Supplementary Data

Supplementary materials are available at *Open Forum Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

### Acknowledgment

**Potential conflicts of interest.** All authors: No potential conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

### References

1. Goswami ND, Pfeiffer CD, Horton JR, et al. The state of infectious diseases clinical trials: a systematic review of ClinicalTrials.gov. *PLoS One* **2013**; 8:e77086.
2. Johnson L, Stricker RB. The Infectious Diseases Society of America Lyme guidelines: a cautionary tale about the development of clinical practice guidelines. *Philos Ethics Humanit Med* **2010**; 5:9.
3. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals: obligation to register clinical trials. Available at: [http://www.icmje.org/publishing\\_10register.html](http://www.icmje.org/publishing_10register.html). Accessed 12 September 2018.
4. Tse T, Fain KM, Zarin DA. How to avoid common problems when using ClinicalTrials.gov in research: 10 issues to consider. *BMJ* **2018**; 361:k1452.
5. Califf RM, Zarin DA, Kramer JM, et al. Characteristics of clinical trials registered in ClinicalTrials.gov, 2007–2010. *JAMA* **2012**; 307:1838–47.
6. Goswami ND, Tsalik EL, Naggie S, et al. A cross-sectional analysis of HIV and hepatitis C clinical trials 2007 to 2010: the relationship between industry sponsorship and randomized study design. *Trials* **2014**; 15:31.
7. Ruopp M, Chiswell K, Thaden JT, et al. Respiratory tract infection clinical trials from 2007 to 2012. A systematic review of ClinicalTrials.gov. *Ann Am Thorac Soc* **2015**; 12:1852–63.

8. Thaden JT, Chiswell K, Jaffe I, et al. pediatric antibacterial and antifungal trials from 2007 to 2017. *Pediatrics*. **2018**; 142.
9. Food and Drug Administration. Food and Drug Administration Amendments Act (FDAAA) of 2007. **2018**. Available at: <https://www.fda.gov/regulatoryinformation/lawsenforcedbyfda/significantamendmentstothefdact/foodanddrugadministrationamendmentsof2007/default.htm>. Accessed 17 September 2018.
10. Clinical Trials Transformation Initiative. Welcome to CTTI. Available at: <https://aact.ctti-clinicaltrials.org/>. Accessed 15 September 2018.
11. 2016 Medical subject headings (MeSH). **2016**. Available at: [https://www.nlm.nih.gov/pubs/techbull/nd15/nd15\\_2016\\_mesh\\_avail.html](https://www.nlm.nih.gov/pubs/techbull/nd15/nd15_2016_mesh_avail.html). Accessed 20 September 2018.
12. Tasneem A, Abele L, Anent H, et al. The database for aggregate analysis of ClinicalTrials.gov (AACT) and subsequent regrouping by clinical specialty. *PLoS One* **2012**; 7:e33677.
13. World Health Organization. Global health estimates 2015: burden of disease by cause, age, sex, by country and by region, 2000–2015. Geneva, Switzerland: World Health Organization; **2016**.
14. World Health Organization. Global health estimates 2015: deaths by cause, age, sex, by country and by region, 2000–2015. Geneva, Switzerland: World Health Organization; **2016**.
15. ClinicalTrials.gov. Map of all studies on ClinicalTrials.gov. Available at: <https://www.clinicaltrials.gov/ct2/search/map>. Accessed 16 September 2018.
16. Pawlotsky J-M, Aghemo A, Back D, et al. EASL recommendations on treatment of hepatitis C 2015. *J Hepatol* **2015**; 63:199–236.
17. ClinicalTrials.gov. A single ascending dose study of daclatasvir (BMS-790052) in hepatitis C virus infected subjects. **2007**. ClinicalTrials.gov identifier: NCT00546715. Available at: <https://clinicaltrials.gov/ct2/show/study/NCT00546715>. Accessed 16 September 2018.
18. ClinicalTrials.gov. Dose-ranging study of sofosbuvir in combination with pegylated interferon and ribavirin in treatment naïve GT 1 HCV patients. **2010**; <https://clinicaltrials.gov/ct2/show/NCT01054729>. Accessed 18 September 2018.
19. ClinicalTrials.gov. A study of TMC435350 administered with or without standard of care therapy in participants with genotype 1 hepatitis C virus infection. **2007**; <https://clinicaltrials.gov/ct2/show/NCT00561353>. Accessed 18 September 2018.
20. Chhatwal J, Kanwal F, Roberts MS, Dunn MA. Cost-effectiveness and budget impact of hepatitis C virus treatment with sofosbuvir and ledipasvir in the United States. *Ann Intern Med* **2015**; 162:397–406.
21. Becker JE, Krumholz HM, Ben-Josef G, Ross JS. Reporting of results in ClinicalTrials.gov and high-impact journals. *JAMA* **2014**; 311:1063–5.
22. Dwan K, Altman DG, Clarke M, et al. Evidence for the selective reporting of analyses and discrepancies in clinical trials: a systematic review of cohort studies of clinical trials. *PLoS Med* **2014**; 11:e1001666.
23. Barker AD, Sigman CC, Kelloff GJ, et al. I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin Pharmacol Ther* **2009**; 86:97–100.
24. Neumann G, Noda T, Kawaoka Y. Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature* **2009**; 459:931–9.
25. Benton A, Dionne KY. International political economy and the 2014 West African Ebola outbreak. *African Stud Rev* **2015**; 58:223–236.
26. World Health Organization. Zika situation report: Zika and potential complications. Geneva, Switzerland: World Health Organization; **2016**.
27. Gill CJ. How often do US-based human subjects research studies register on time, and how often do they post their results? a statistical analysis of the ClinicalTrials.gov database. *BMJ Open* **2012**; 2: e001186.