

Trends of Clinical Trials for Drug Development in Rare Diseases

Ryuichi Sakate¹, Akiko Fukagawa², Yuri Takagaki², Hanayuki Okura¹ and Akifumi Matsuyama^{1,2,†,*}¹Center for Rare Disease Research, National Institutes of Biomedical Innovation, Health and Nutrition, Osaka, Japan;²Office of Policy and Ethics Research, National Institutes of Biomedical Innovation, Health and Nutrition, Osaka, Japan

Abstract: Background: Drug development for rare diseases is challenging because it is difficult to obtain relevant data from very few patients. It must be informative to grasp current status of clinical trials for drug development in rare diseases.

Objective: Clinical trials in rare diseases are to be outlined and compared among the US, EU and Japan.

Method: ClinicalTrials.gov (NCT, National Clinical Trial), EU Clinical Trials Register (EUCTR) and the Japan Primary Registries Network (JPRN) were analyzed. Clinical trials involving information on rare diseases and drugs were extracted by text-mining, based on the diseases and drugs derived from Orphanet and DrugBank, respectively.

Results: In total, 28,526 clinical trials were extracted, which studied 1,535 rare diseases and 1,539 drugs. NCT had the largest number of trials, involving 1,252 diseases and 1,332 drugs. EUCR and JPRN also had registry-specific diseases (250 and 22, respectively) and drugs (172 and 29, respectively) that should not be missed. Among the 1,535 rare diseases, most diseases were studied in only a limited number of trials; 70% of diseases were studied in fewer than 10 trials, and 28% were studied in only one. Additionally, most studied rare diseases were cancer-related ones.

Conclusion: This study has revealed the characteristics of the clinical trials in rare diseases among the US, EU and Japan. The number of trials for rare diseases was limited especially for non-cancer-related ones. This information could contribute to drug development such as drug-repositioning in rare diseases.

Keywords: ClinicalTrials.gov, EU Clinical Trials Register, the Japan Primary Registries Network, Orphanet, DrugBank, a limited number of trials, cancer-related rare diseases, drug repositioning.

1. INTRODUCTION

A rare disease is defined as a disease that affects a small number of people. It was reported that the number of people affected by rare diseases was $\leq 5/10,000$ in Europe, $< 7/10,000$ in the United States and $< 4/10,000$ in Japan [1, 2]. However, there are approximately 7,000 rare diseases in total and more than 300 million people affected worldwide [3]. Because of the low prevalence of each rare disease, they are difficult to diagnose, and treatment remains challenging. Many of these diseases are severe and involve life-long suffering. The unmet medical needs for rare disease treatment are growing worldwide.

In our previous study, we focused on 306 intractable diseases (mostly rare diseases) designated by the Ministry of Health, Labour and Welfare in Japan, and clinical trials for drug development were analyzed [4-8]. As a result, more than 4,000 trials, involving 152 diseases and 577 drugs, were found. It was noted that trials were conducted only for approximately half of the diseases (152/306). What interested us most was that drugs were shared among a considerable number of diseases (for example, Cyclophosphamide was used in the trials for 30 diseases). This information could be particularly valuable for drug-repositioning. An inter-disease comparison should be one of the most important factors in drug development, especially for rare diseases that have less information.

This study aimed to expand the target rare diseases from 306 to all as much as possible and to clarify worldwide trends in drug development. As a more comprehensive disease set, rare diseases registered in Orphanet [9] were selected. Orphanet contained 9,525 rare disease names. It included 290 of the 306 diseases of our previous study [6]. As for comprehensive clinical trial data, ClinicalTrials.gov

*Address correspondence to this author at the Center for Rare Disease Research, National Institutes of Biomedical Innovation, Health and Nutrition, 7-6-8 Saito-Asagi, Ibaraki-shi, Osaka 567-0085, Japan; Tel/Fax: +81-72-641-9016; E-mail: akifumi-matsuyama@nibiohn.go.jp

†Present address: Department of Regenerative Medicine and Stem Cell Biology, School of Medicine, Fujita Health University, 1-98 Dengakugakubo, Kutsukake-cho, Toyoake, Aichi 470-1192, Japan; Tel: +81-562-93-2000; E-mail: akifumi-matsuyama@umin.ac.jp

(NCT, National Clinical Trial) [10], EU Clinical Trials Register (EUCTR) [11], and the Japan Primary Registries Network (JPRN) [12], representing the United States of America and European Union (EU) and Japan, respectively, were selected. These registries were the top 3 in the number of trials in the WHO International Clinical Trials Registry Platform (WHO ICTRP) [13].

Other studies analyzing clinical trials in rare diseases have produced useful information. One study compared rare versus non-rare diseases [14]. Another assessed the availability of the results of phase 3 and 4 trials [15]. A third study analyzed the correlation between low prevalence and sample size [16]. While these three studies used NCT data, the present study used EUCTR and JPRN, as well as NCT, and clarified the similarities and differences among them. In this way, a more comprehensive analysis of clinical trials was conducted.

The objective of this study is to outline and compare clinical trials in rare diseases among the US, EU and Japan, revealing the relationship between diseases and drugs as well as the number of trials, diseases and drugs. The number of drugs sharing a disease was analyzed, and vice versa. This information will be considerably useful for drug development.

Technically, an underlying theme of this study is “Mining and connecting data.” Different formats in different databases were mediated and connected in this study. This was for the data among the registries, trials, diseases and drugs. While making maximum use of available databases, connecting unconnected data would be indispensable as we move towards the era of big data.

2. MATERIAL AND METHOD

Searches of WHO ICTRP with “Recruitment status: ALL” option were conducted and all clinical trial data were downloaded. From the 357,384 trials downloaded, trials of the three registries (ClinicalTrials.gov, EUCTR and JPRN) were extracted by distinguishing their id prefixes (NCT, EUCTR and JPRN). The number of trials became 288,217 (NCT: 234,486, EUCTR: 26,153 and JPRN: 27,578). As for controlled rare disease names and drug names, Orphadata data [17] (Orphanet data repository; en_product1.xml.gz) and DrugBank [18, 19] (drugbank_all_full_database.xml.zip, version 5.0.6) were downloaded, respectively.

Firstly, trials involving drug information were extracted by searching for specific keywords (*for EUCTR: "Trade Name:," "Product Name:," "INN or Proposed INN:," "Product Code:," "Other descriptive name:," and for NCT: "Drug:," "Biological:," "Dietary Supplement:"*) in the “Intervention” section of the trial data. Words and phrases following these keywords were extracted. Then, 8,283 drugs from DrugBank (name, synonyms, products and international brands for each drug) were searched in the extracted words and phrases and, for JPRN, in the whole “Intervention” section. The trials involving at least one drug were extracted. Less than four-letter drug names or less specific ones were eliminated (*Alert, Beam, Body, Care, Contingency, Cough, Easy, Equal, Fast, Hand, Healing, Leader, Maxim, Muscle, Nasal, Night Time, Patch, Prevent, Protection, Purpose,*

Purpose, Recovery, Regular, Renewal, SAME, Screen, Sleep, Spectrum, Success, Solution, Swab, Therapeutic, Water). Then, the number of trials became 162,318 (NCT: 127,025, EUCTR: 26,013 and JPRN: 9,280).

Secondly, trials involving rare disease names were extracted by searching for 9,525 disease names of Orphanet (name and synonyms for each disease), word by word, in “Condition” section of the trial data. The trials involving at least one disease were extracted. Less than four-letter disease names (abbreviations) or less-specific words/symbols were eliminated (*a, above, across, after, an, around, as, at, and, before, behind, below, beside, between, but, by, down, due, during, for, from, had, has, have, in, INCL, inside, is, it, its, made, may, of, on, onto, or, other, RICH, some, that, the, their, they, through, to, toward, towards, under, up, us, was, we, were, which, with*). Additional adjustments were conducted for plural/singular forms, special characters (*À, Ê, á, ä, ç, è, é, ë, í, ï, ñ, ó, ö, ü*), synonyms (*disease/disorder/syndrome, cancer/carcinoma*) and others (*to distinguish non-Hodgkin lymphoma and Hodgkin lymphoma, and non-small cell lung cancer and small cell lung cancer*). Finally, the number of trials became 28,526 (NCT: 21,981, EUCTR: 5,363 and JPRN: 1,182).

Properties attached to the 28,526 trials were summarized for the three registries (recruitment status, age, gender, target size, phase and countries). The number of trials, diseases and drugs were analyzed in combination with each other by our in-house programs.

This study used publicly available data of clinical trials and focused on their characteristics rather than human participants themselves. Thus, ethical committee approval was not required.

3. RESULTS

3.1. Characteristics of the Three Registries

In total, 28,526 clinical trials were extracted from the three registries (NCT, EUCTR and JPRN) in WHO ICTRP. They contained information on 1,535 rare diseases and 1,539 drugs. The characteristics of the three registries are summarized in Table 1. 21,981 trials were registered in NCT (as much as 77.1% of all trials). The number of diseases (1,252) and drugs (1,332) were also large in NCT. Smaller numbers of trials were registered in EUCTR and JPRN (5,363 and 1,182, respectively). However, the data contained a considerable number of diseases (957 and 295, respectively) and drugs (991 and 389, respectively). For example, while NCT was nearly as much as 20 times larger than JPRN in the number of trials, the difference became smaller in the numbers of diseases and drugs (NCT was 3–4 times larger than JPRN).

Registry-specific diseases and drugs were also found (Fig. 1). While 517 diseases were specific to NCT, 250 diseases (IMAGE syndrome, Kabuki syndrome, *etc.*) and 22 diseases (Budd-Chiari syndrome, Isaac syndrome, *etc.*) were specific to EUCTR and JPRN, respectively (Fig. 1B). For drugs, while 460 were specific to NCT, 172 drugs (Rizatriptan, Peginesatide, *etc.*) and 29 drugs (Benidipine, Imidafenacin, *etc.*) were specific to EUCTR and JPRN, respectively (Fig. 1C).

Table 1. Characteristics of the NCT, EUCTR and JPRN trials.

		NCT		EUCTR		JPRN		Total/(%)		
#Trial		21,981	(77.1)	5,363	(18.8)	1,182	(4.1)	28,526	(100.0)	
Registration year		1999-2017		2004-2017		2005-2017		1999-2017		
#Disease*		1,252	(81.6)	957	(62.3)	295	(19.2)	1,535	(100.0)	
#Drug*		1,332	(86.5)	991	(64.4)	389	(25.3)	1,539	(100.0)	
Recruitment status	Recruiting	5,422	(24.7)	-	-	558	(47.2)	5,980	(21.0)	
	Authorized	-	-	3,020	(57.3)	-	-	3,020	(10.6)	
	Not recruiting	16,559	(75.3)	2,251	(42.7)	624	(52.8)	19,434	(68.3)	
	N/A	-	-	92	-	-	-	92	-	
Age	Pediatric (<18)	Yes	4,301	(21.8)	-	-	218	(20.5)	4,519	(21.7)
		No	15,470	(78.2)	-	-	846	(79.5)	16,316	(78.3)
	Elderly (>65)	Yes	3,768	(39.6)	-	-	458	(73.5)	4,226	(41.6)
		No	5,758	(60.4)	-	-	165	(26.5)	5,923	(58.4)
	N/A	1,077	-	5,363	-	95	-	6,535	-	
Gender	Female	1,328	(6.0)	315	(5.9)	59	(5.0)	1,702	(6.0)	
	Male	550	(2.5)	172	(3.2)	10	(0.8)	732	(2.6)	
	Both	20,101	(91.5)	4,871	(90.9)	1,113	(94.2)	26,085	(91.5)	
	N/A	2	-	5	-	0	-	7	-	
Target size	Average	456		244		67		409		
	1-50	11,492	(55.8)	1,021	(30.5)	832	(70.4)	13,345	(53.2)	
	51-100	3,885	(18.9)	695	(20.8)	206	(17.4)	4,786	(19.1)	
	101-500	4,104	(19.9)	1,254	(37.5)	127	(10.7)	5,485	(21.8)	
	501-	1,099	(5.3)	374	(11.2)	17	(1.4)	1,490	(5.9)	
	N/A	1,401	-	2,019	-	0	-	3,420	-	
Phase	Phase 1	6,555	(29.5)	151	(11.2)	140	(20.6)	6,846	(28.3)	
	Phase 2	10,735	(48.3)	675	(50.3)	441	(64.8)	11,851	(48.9)	
	Phase 3	3,574	(16.1)	382	(28.4)	86	(12.6)	4,042	(16.7)	
	Phase 4	1,342	(6.0)	135	(10.1)	14	(2.1)	1,491	(6.2)	
	N/A	2,621	-	4,174	-	607	-	7,402	-	
	Total**	22,206	(100.0)	1,343	(100.0)	681	(100.0)	24,230	(100.0)	
Countries	#Country*	161	(96.4)	103	(61.7)	6	(3.6)	167	(100.0)	
	Top 5 (#Trial)	1	13,006	United States	1,820	Germany	1,178	Japan	13,812	United States
		2	1,889	Canada	1,688	United Kingdom	11	Asia(except Japan)	3,312	Germany
		3	1,864	France	1,640	Italy	4	Europe	3,177	United Kingdom
		4	1,492	Germany	1,233	Spain	3	North America	3,014	France
		5	1,489	United Kingdom	1,150	France	3	Australia	2,911	Italy

*Diseases, drugs or countries could be shared by two or three registries.

**Redundancy included. Cases where more than one Phase per trial exist (ex. Phase 1/2 in one trial).

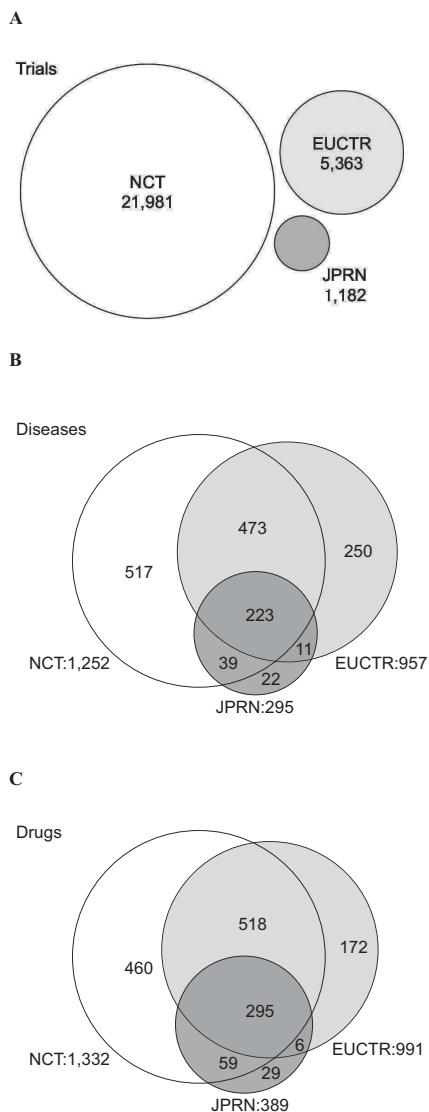


Fig. (1). Comparison of the three registries (NCT, EUCTR, and JPRN). **(A)** Number of trials. **(B)** Number of diseases. **(C)** Number of drugs. Shared and registry-specific states for diseases and drugs were shown in the Venn diagrams.

The numbers of diseases and drugs shared by all the three registries were 223 and 295, respectively (Fig. 1B and 1C).

Other characteristics were discovered among the registries as well. The following sections provide a brief overview of the characteristics such as recruitment status, age, gender, target size, phase and countries (Table 1).

“Not recruiting” trials occupied a larger proportion (75.3%) in NCT than in EUCTR (42.7%) and JPRN (52.8%). This might be because older trials were more abundant in NCT (registration since 1999) than in EUCTR (since 2004) and JPRN (since 2005).

As shown in age data, the inclusion of elderly people (> 65 years old) was nearly twice as large in JPRN (73.5%) than in NCT (39.6%) (no age data obtained in EUCTR). Elderly people specifically attended the trials of JPRN. In gender data, more trials for females than males were found con-

sistently among the three registries. This might be because more female-specific diseases (Ovarian cancer, *etc.*) were registered in Orphanet.

In terms of target size, larger scale trials were conducted in NCT, as shown in the average size of NCT (456 people), EUCTR (244) and JPRN (67). It should be noted that trials with > 100 people (Target size: 101-500, 501-) occupied the largest part (48.7%) in EUCTR, while they were 25.3% in NCT and 12.2% in JPRN, respectively. Trials with ≤ 50 people (Target size: 1-50) were the most common in NCT (55.8%) and JPRN (70.4%). A small number of trials with very large target size made the average higher in NCT.

In phase data, there was strong similarity among the three registries that nearly a half or more trials were registered as Phase 2 (NCT: 48.3%, EUCTR: 50.3% and JPRN: 64.8%).

To view regional characteristics of the three registries, the top 5 countries where the trials were conducted were compared. Of the 21,981 trials of NCT, 13,006 trials (59.2%) were conducted in the United States. Germany, the United Kingdom and France were listed in the top 5 of EUCTR and NCT trials. This showed that many trials were conducted

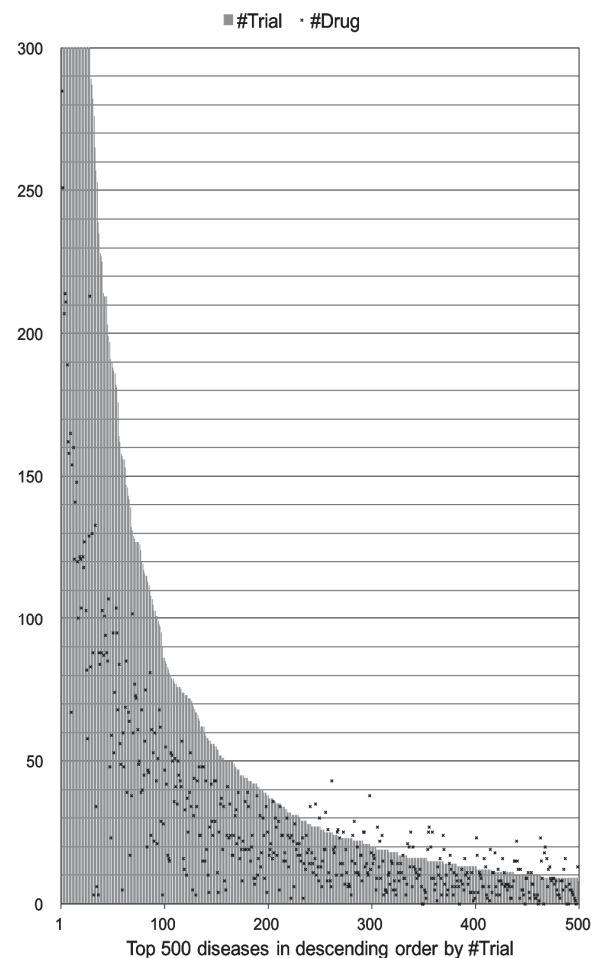


Fig. (2). Distribution of the numbers of trials and drugs. The top 500 diseases were shown in descending order by the number of trials. A long-tailed distribution was found.

internationally between the United States and Europe. However, in JPRN, almost all trials were conducted only in Japan (1,178, 99.7%).

3.2. Relationships among Clinical Trials, Diseases and Drugs

When analyzing all trials together, a large number of trials were conducted for only a small number of diseases (Fig. 2). While six diseases (0.4% of the 1,535 diseases) had > 1,000 trials (Table 2), 69.8% of diseases had < 10 trials, and 27.9% had only one trial. For each disease, the number of trials ranged from 1 to 4,084 (a list of the 1,535 diseases is available in Supplementary Table 1). The relationship between trials and drugs for each disease seemed to vary, as shown in Fig. 2. For one disease, a small number of drugs were tested in many trials. For another disease, the number of drugs tested was larger than the number of trials. The latter case suggests that two or more drugs were tested per trial frequently.

The top 20 most studied diseases in the trials are listed in Table 2. The top 3 diseases were Lymphoma (Orpha number: 223735), Multiple myeloma (29073), and Ovarian can-

cer (213500). Lymphoma was the most studied disease in NCT and EUCTR, and the third in JPRN (Supplementary Table 2). It seemed that cancer-related rare diseases were the most prevalent on the list. In fact, the number of diseases with their names including "blastoma," "cancer," "carcinoma," "leukemia," "lymphoma," "myeloma" or "tumor" was only 247 (16.1%) of the 1,535 diseases. This suggested that a large number of trials were conducted for a relatively small number of cancer-related rare diseases.

The top 20 most tested drugs in descending order by the number of rare diseases are listed in Table 3. The top 3 drugs were Cyclophosphamide (DrugBank Accession Number: DB00531), Cyclosporine (DB00091) and Methotrexate (DB00563). Cyclophosphamide was shared by 261 diseases. It was also the most shared drug in all the three registries (Supplementary Table 3). This drug was also the top in our previous study targeting the 306 intractable diseases (mostly rare diseases) in Japan [4]. As expected from the top 20 diseases (Table 2), antineoplastic drugs were abundant in the list. In addition, immunosuppressive drugs or anti-inflammatory drug were found in abundance. Of the 1,539 drugs, 29 drugs (1.9%) were shared by > 100 diseases, 387 drugs (25.1%)

Table 2. The top 20 most studied rare diseases in descending order by the number of trials.

	Orpha num	Name	Type	ICD-10	#Trial	#Drug
1	223735	Lymphoma	Group of phenomes		4,084	417
2	29073	Multiple myeloma	Disease	C90.0	1,879	285
3	213500	Ovarian cancer	Group of phenomes		1,381	251
4	519	Acute myeloid leukemia	Group of phenomes	C92.0	1,297	207
5	52688	Myelodysplastic syndrome	Group of phenomes		1,243	214
6	547	Non-Hodgkin lymphoma	Group of phenomes		1,081	211
7	88673	Hepatocellular carcinoma	Disease	C22.0	919	189
8	217071	Renal cell carcinoma	Group of phenomes		813	162
9	513	Acute lymphoblastic leukemia	Group of phenomes	C91.0	758	158
10	360	Glioblastoma	Disease	C71.9	745	165
11	673	Malaria	Disease	B51.0	674	67
12	586	Cystic fibrosis	Disease	E84.9	662	154
13	70482	Carcinoma of esophagus	Group of phenomes		635	160
14	521	Chronic myeloid leukemia	Disease	C92.1	601	121
15	98293	Hodgkin lymphoma	Group of phenomes		550	141
16	182067	Glial tumor	Group of phenomes		498	148
17	545	Follicular lymphoma	Disease	C82.0	474	120
18	3389	Tuberculosis	Disease		451	100
19	180242	Malignant tumor of fallopian tubes	Disease	C57.0	450	122
20	52416	Mantle cell lymphoma	Disease	C83.1	428	121

Table 3. The top 20 most tested drugs in descending order by the number of diseases.

	Acc Num	Name	Type	Groups*	#Disease
1	DB00531	Cyclophosphamide	small molecule	approved investigational	261
2	DB00091	Cyclosporine	small molecule	approved investigational vet_approved	192
3	DB00563	Methotrexate	small molecule	approved	178
4	DB00773	Etoposide	small molecule	approved	172
5	DB00099	Filgrastim	biotech	approved	171
6	DB00112	Bevacizumab	biotech	approved investigational	170
7	DB01234	Dexamethasone	small molecule	approved investigational vet_approved	162
8	DB01073	Fludarabine	small molecule	approved	159
9	DB01008	Busulfan	small molecule	approved investigational	158
10	DB00688	Mycophenolate mofetil	small molecule	approved investigational	152
11	DB01024	Mycophenolic acid	small molecule	approved	151
12	DB00958	Carboplatin	small molecule	approved	145
13	DB00997	Doxorubicin	small molecule	approved investigational	142
14	DB01042	Melphalan	small molecule	approved	139
15	DB00959	Methylprednisolone	small molecule	approved vet_approved	132
16	DB00635	Prednisone	small molecule	approved vet_approved	129
17	DB00188	Bortezomib	small molecule	approved investigational	127
17	DB00073	Rituximab	biotech	approved	127
19	DB00877	Sunitinib	small molecule	approved investigational	121
20	DB02546	Vorinostat	small molecule	approved investigational	118
20	DB00864	Tacrolimus	small molecule	approved investigational	118

*approved: approved in at least one jurisdiction, at some point in time.

vet_approved: approved in at least one jurisdiction, at some point in time for the treatment of animals.

investigational: in some phase of the drug approval process in at least one jurisdiction.

<https://www.drugbank.ca/documentation>

were shared by > 10 diseases, and 401 drugs (26.1%) were shared by only one disease.

The number of trials conducted per drug was also counted (Supplementary Table 4). Cyclophosphamide (DB00531) was the most tested drug by as many as 1,396 trials in total. While Cyclophosphamide was also the most tested drug in NCT, Rituximab (DB00073) and Cisplatin (DB00515) were the most tested drugs in EUCR (235 trials) and JPRN (147 trials), respectively.

3.3. Progress of the Clinical Trials

For the 1,535 diseases, the trials have been registered since 1999. The first registration year of the trials for each disease is shown in Fig. 3A. From 1999–2016, 84 diseases per year on average were newly targeted. There were two peaks in 1999 (196 diseases) and 2005 (187 diseases). The 1999 peak could be explained by the fact that registrations of earlier trials were included at the start of the registry. The 2005 peak could be explained by the fact that there had been

a worldwide push from governments to make registration mandatory. After the 2005 peak, the number of newly registered diseases decreased. Recently (2010–2016), they were 50–80 diseases per year.

Of the 28,526 trials, 21,124 trials (74.1%) involved phase information (Phase 1–4) (Fig. 3B). The registration year of the trials rather resembled the cumulative graph in Fig. 3A, showing that the number of trials increased as the number of diseases increased. The peaks of 1999 (711 trials in total) and 2005 (1,454 trials in total) were vague. The number of trials was gradually growing since 2005, and the pace became faster in 2014–2016. It should be noted that the number of Phase 2 trials was the largest consistently during all the period, and that the number of trials was large in order of Phase 2, Phase 1, Phase 3 and Phase 4 for most of the period.

In order to assess the progress of the trials, the Phase 3 and Phase 4 trials were selected as promising trials. For each disease, the numbers of Phase 3 and Phase 4 trials were calculated and their proportion to all trials (Phase 1–4) was

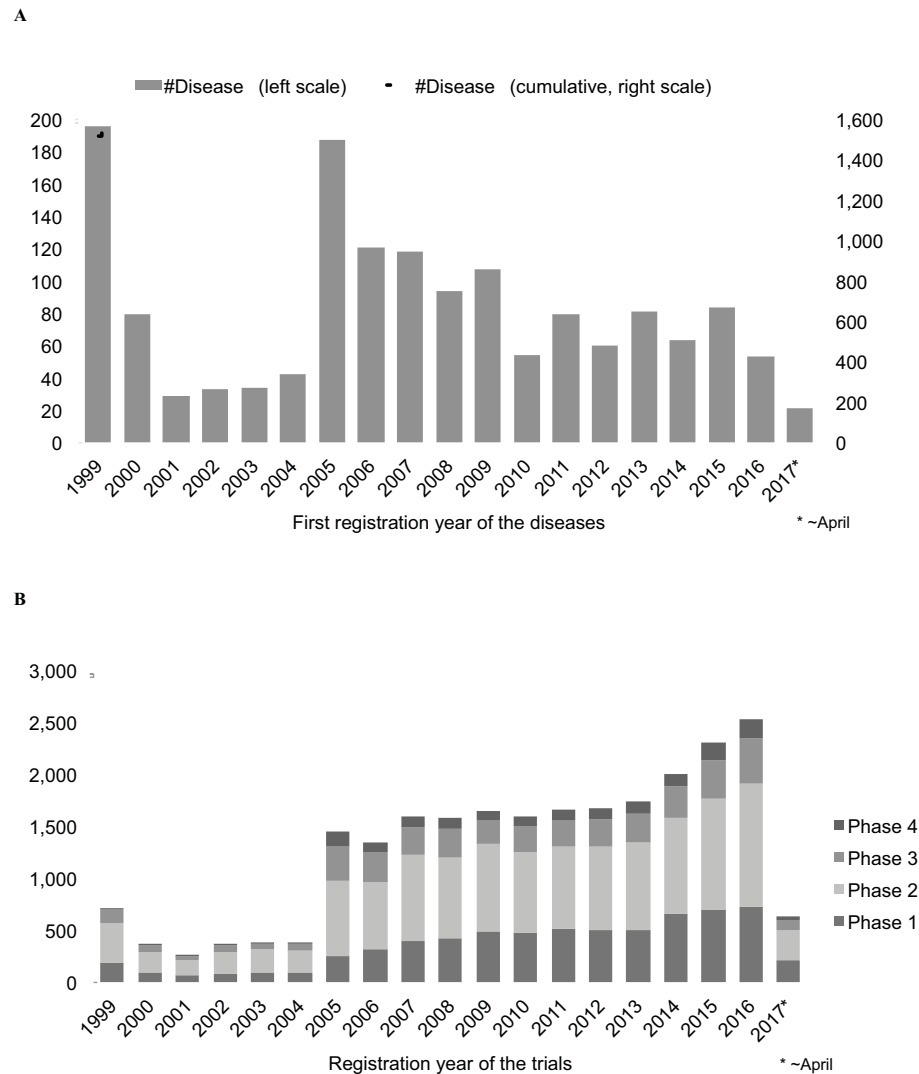


Fig. (3). Registration years of the trials. **(A)** First registration year of the trials for each disease. **(B)** Registration year of the trials and their phases (21,251 trials with phase information). Cases where more than one phase existed per trial were multi-counted (for example, in a case where Phase 1 and Phase 2 existed per one trial, two trials were counted.).

compared. For 1,241 diseases with trials involving phase information, the distribution of the proportions (0%, 0–25%, 25–50%, 50–75% and 75–100%) was shown in Fig. 4A. More than a half of the diseases (735 diseases, 59.2%) had the proportion of Phase 3 and Phase 4 trials equal or less than 25% (0%, 0–25%). Furthermore, diseases with neither Phase 3 nor Phase 4 trials (0%) occupied as much as 37.1% (460 diseases). Even if the number of trials was large, the proportion of Phase 3 and Phase 4 trials tended to be small for most diseases (Fig. 4B). This means that most trials were interrupted in Phase 1 or Phase 2 and could not proceed to Phase 3 and Phase 4. The difficult situation in drug development for rare diseases might be inferred.

4. DISCUSSION

4.1. Characteristics of the Three Registries and Disease-Drug Relationship

In this study, clinical trials, rare diseases and drugs were successfully connected with each other by our original text

mining analyses. As a result, (1) characteristics of the three registries (NCT, EUCTR and JPRN) and (2) disease-drug relationship in the clinical trials were clarified.

Among the three registries, NCT appeared to be the largest registry of clinical trials also in rare diseases. In addition, many trials conducted in Europe were also registered in NCT (Table 1). The centralized information of NCT might reflect a powerful driving force of drug development in the United States, which is one of the largest pharmaceutical markets in the world.

However, differences in characteristics were observed among the three registries. For example, regarding the inclusion age, a proportion of the number of trials with elderly people (> 65 years old) was specifically higher in JPRN (73.5%) than NCT (39.6%). Regarding target size, a proportion of the number of trials with > 100 people was uniquely higher in EUCTR (48.7%) than in NCT (25.3%) and JPRN (12.2%), though the average size was smaller in EUCTR (244) than in NCT (456).

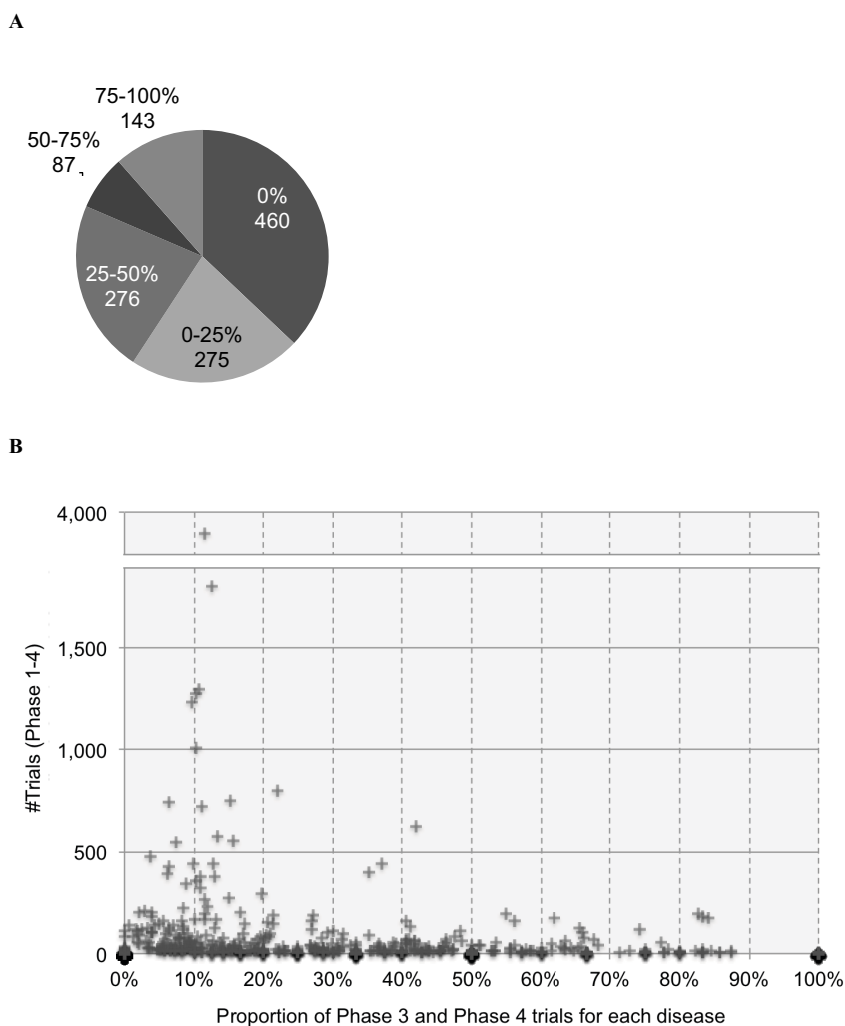


Fig. (4). Proportion of Phase 3 and Phase 4 trials to all trials (Phase 1–4) for each disease. **(A)** Distribution of the proportion (%) of Phase 3 and Phase 4 trials for each disease (in total, 1,241 diseases with phase information). For example, the number of diseases with 0% of Phase 3 and Phase 4 trials (i.e., no Phase 3 and Phase 4 trials) was 460. **(B)** Scatter plot of the 1,241 diseases with their number of trials (y-axis) and their proportion of Phase 3 and Phase 4 trials (x-axis).

NCT had the largest number of trials, involving 1,252 diseases and 1,332 drugs. However, in EUCTR and JPRN, a considerable number of registry-specific diseases (250 and 22, respectively) and drugs (172 and 29, respectively) were found (Fig. 1). For example, a trial for Budd-Chiari syndrome (Orpha number: 131) was registered only in JPRN. In addition, trials for Kabuki syndrome (Orpha number: 2322) were registered only in EUCTR. Previously a region-specific prevalence was reported for the Behçet disease (Orpha number: 117) [20]. As revealing region-specificity in rare diseases needs further investigation, a broad collection of clinical trials data worldwide would be advantageous.

In this study, information on trials, diseases and drugs were analyzed by counting (i) the number of trials conducted per disease (Table 2 and Supplementary Table 2), (ii) the number of diseases studied per drug (Table 3 and Supplementary Table 3) and (iii) the number of trials conducted per drug (Supplementary Table 4). This information could promote investigations of (i) what trials were conducted for what diseases, (ii) what drugs were shared by what diseases

and (iii) what trials were conducted for what drugs. These investigations will be expanded by further analyses such as tracing a drug to be launched, comparing target genes among drugs, searching for drug repositioning targets.

Trials and drugs were found to be limited to only a small number of rare diseases. Interestingly and essentially, the top 20 diseases were mostly related to rare cancers. Although the number of these cancer-related rare diseases was limited to 247 (16% of the 1,535 rare diseases in this study), as many as 14,594 trials (51% of the 28,576 trials) were conducted and 800 drugs (52% of the 1,539 drugs) were tested for the diseases. Thus, it was understandable that the most tested drugs were antineoplastic drugs, immunosuppressive drugs and anti-inflammatory drugs that were often used for cancers. This resulted in that, in fact, even less trials were conducted for non-cancer-related (ordinary, so to say) rare diseases, though a border between “cancer-related” and “non-cancer-related” rare diseases might be in a grey-zone in the definition.

While Phase 3 and Phase 4 trials were considered as promising for drug development, a small proportion of the

trials seemed to reach Phase 3. For 460 diseases (37.1%), Phase 3 and Phase 4 trials were not conducted at all. This trend appeared even in diseases for which many trials were conducted (Fig. 4B). This suggested that, currently, most of the trials in rare diseases were not promising in drug development worldwide. Further analyses could investigate this trend comparing among rare diseases and other disease groups.

As described above, only a small number of trials were conducted in rare diseases (non-cancer-related rare diseases). For example, the number of trials conducted for Amyotrophic lateral sclerosis (ALS) (Orpha number: 803) was 227, and it was far removed from the top 20 diseases (Table 2). Moreover, for ALS, the number of Phase 3 and Phase 4 trials accounted for only 16.7% (34 out of 203 trials involving phase information). The clinical trials for drug development in rare diseases were found to be just at the beginning stages of their progress.

4.2. Limitation of the Analyses and Future Perspectives

Orphanet was known to be one of the most comprehensive databases for rare diseases. In this study, of the 9,525 diseases of Orphanet (including groups of diseases), 1,535 diseases (16.1%) were found in the clinical trials for drug development (Table 1). This number might be less than expected and reflect difficulties in conducting clinical trials in rare diseases (low prevalence, incomplete diagnosis, *etc.*). These difficult situations might be improved by further integration among databases for rare disease. For example, in this study, we had to search for rare disease names in the trial data by text mining. Currently, this is the case for rare diseases where the data were not yet connected or incompletely connected among databases. The assignment of ICD-10 [21] and other terms and ontology to rare diseases seemed to be insufficient. Thus, constructing annotated and standardized data of rare diseases is strongly required in further studies for drug development.

WHO ICTRP integrated all clinical trials in the world (not limited to rare diseases). Other than NCT, EUCTR and JPRN, many registries were registered in WHO ICTRP. In particular, the number of trials conducted by Chinese institutions was found to be increasing [22, 23]. Also, the number of trials in the Chinese Clinical Trial Registry (ChiCTR) was growing rapidly. In this regard, more inter-registry integration of the clinical trial data will be valuable especially for analyzing regionality and overcoming low prevalence in rare diseases.

The limitations on the analyses of rare diseases were derived from their nature. The definition and standardization of rare diseases were essentially based on long-standing medical and scientific studies. These studies need more time to become sufficient for drug development studies. In addition, integration and standardization for precise and up-to-date information of rare diseases, including clinical trials, could only be established through international collaboration. For example, the International Rare Diseases Research Consortium (IRDiRC) [24] was expected to contribute largely to these objectives.

CONCLUSION

In this study, drug development in rare diseases was first assessed comprehensively by extracting 28,526 clinical trials from the three registries (NCT, EUCTR and JPRN). While NCT played a significant role, the information obtained from other registries (EUCTR and JPRN) were also indispensable because they contained trials for specific rare diseases that were not found elsewhere.

Overall, trials and drugs were limited to a small number of rare diseases. These diseases were often cancer-related rare diseases. For most of the rare diseases (non-cancer-related ones), as a result, the numbers of trials and drugs were very small. In addition, considering a smaller proportion of their Phase 3 and Phase 4 trials, drug development for rare diseases appeared to have a higher hurdle to clear.

Integrating the information of trials, diseases and drugs is expected to boost investigation in rare disease drug development. The integration could contribute to designing more efficient clinical trials by searching drug-repositioning candidates. Also, it could contribute to new findings of disease characteristics to define their diagnosis. This study could pose a starting point for a series of further rare disease studies: disease-related genes and phenotypes, pathogenic mechanisms, drug classifications, political and ethical subjects, pharmaceutical strategies, *etc.* Integrated with various data, clinical trial data could serve as core data for the studies.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No Animals/Humans were used for studies that are the basis of this research.

CONSENT FOR PUBLICATION

Not applicable.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

AM initially planned this study. RS and AF investigated the data. RS wrote in-house programs for the analyses. RS, AF and AM curated and discussed the data. RS wrote the manuscript and AF, HO, YT and AM revised the manuscript. All authors approved the final manuscript. We are grateful to Masashi Suzuki, Masanori Tobe, Hiroto Akabane and Masami Morita (Office of Pharmaceutical Industry Research, Japan Pharmaceutical Manufacturers Association) for their valuable comments on this study.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's web site along with the published article.

REFERENCES

- [1] Kawashima Kodama T. Global strategy for rare and intractable diseases. *Rinsho Shinkeigaku* 2013; 23(11): 1283-6.
- [2] Melnikova I. Rare diseases and orphan drugs. *Nat Rev Drug Discov* 2012; 11(4): 267-8.
- [3] List RARE. Global Genes <https://globalgenes.org/rarelist/> [Accessed August 1, 2017].
- [4] Suzuki M, Sakate R, Fukagawa A. Status of implementation of clinical trials on intractable diseases designated in Japan. *OPIR News* 2016; 48: 23-7.
- [5] Suzuki M, Sakate R, Fukagawa A. Implementers of clinical trials for the designated intractable diseases. *OPIR News* 2016; 49: 29-34.
- [6] Suzuki M, Sakate R, Fukagawa A. Connecting information of the designated intractable diseases to databases. *OPIR News* 2016; 49: 35-46.
- [7] Suzuki M, Sakate R, Fukagawa A. Launched drugs used for clinical trials of the designated intractable diseases. *OPIR News* 2017; 50: 39-43.
- [8] Tobe M, Akabane H, Sakate R, Fukagawa A. Modality analysis of the drugs used for clinical trials of the designated intractable diseases. *OPIR News* 2017; 52: 37-42.
- [9] Orphanet <http://www.orpha.net> [Accessed April 30, 2017].
- [10] <https://clinicaltrials.gov> [Accessed July 9, 2017].
- [11] <https://www.clinicaltrialsregister.eu/> [Accessed July 9, 2017].
- [12] <http://www.who.int/ictcp/network/jpm2/en/> [Accessed July 9, 2017].
- [13] World Health Organization <http://apps.who.int/trialsearch/AdvSearch.aspx> [Accessed April 25-26, 2017].
- [14] Bell SA, Tudur Smith C. A comparison of interventional clinical trials in rare versus non-rare diseases: an analysis of ClinicalTrials.gov. *Orphanet J Rare Dis* 2014; 9: 170.
- [15] Dechartres A, Riveros C, Harroch M, Faber T, Ravaud P. Characteristics and Public Availability of Results of Clinical Trials on Rare Diseases Registered at Clinicaltrials.gov. *JAMA Intern Med* 2016; 176(4): 556-8.
- [16] Hee SW, Willis A, Tudur Smith C, *et al.* Does the low prevalence affect the sample size of interventional clinical trials of rare diseases? An analysis of data from the aggregate analysis of clinicaltrials.gov. *Orphanet J Rare Dis* 2017; 12(1): 44.
- [17] Orphadata. <http://www.orphadata.org> [Accessed April 30, 2017].
- [18] Law V, Knox C, Djoumbou Y, *et al.* DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res* 2014; 42(Database issue): D1091-7.
- [19] DrugBank <http://www.drugbank.ca> [Accessed April 1, 2017].
- [20] Mizuki N, Meguro A, Ota M, *et al.* Genome-wide association studies identify IL23R-IL12RB2 and IL10 as Behçet's disease susceptibility loci. *Nat Genet* 2010; 42(8): 703-6.
- [21] International Classification of Diseases, World Health Organization. <http://www.who.int/classifications/icd/en/> [Accessed Oct 25, 2017].
- [22] Yang L, Su C, Lee AM, Bai HX. Focusing on rare diseases in China: are we there yet? *Orphanet J Rare Dis* 2015; 10: 142.
- [23] Cheng A, Xie Z. Challenges in orphan drug development and regulatory policy in China. *Orphanet J Rare Dis* 2017; 12(1): 13.
- [24] International Rare Diseases Research Consortium (IRDiRC) <http://www.irdirc.org> [Accessed July 31, 2017].