Supplementary Materials

Co-funded by
the European Union

OECD

This document should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed are those of the author(s).

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Note by the Republic of Türkiye:

The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Türkiye recognizes the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of United Nations, Türkiye shall preserve its position concerning the "Cyprus" issue.

Note by all the European Union Member States of the OECD and the European Union:

The Republic of Cyprus is recognised by all members of the United Nations with the exception of Türkiye. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

This document was produced with the financial support of the European Commission - Grant Number 2020 53 03. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the European Union.

© OECD 2023

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for commercial use and translation rights should be submitted to rights@oecd.org.

Table of contents

This document presents material that supplements:

Keelara, R., Wenzl, M., Waagstein, L., Moens, M., Lopert, R. (2023) "Developing a set of indicators to monitor the performance of the pharmaceutical industry", *OECD Health Working Papers*, No. 157, OECD Publishing, Paris, https://doi.org/10.1787/3b5ca61c-en

TABLES

BOXES

Annex A. Data Definitions and Methods

Inputs

1. Indicators of inputs include revenue, net cash flows from financing activities, direct subsidies for R&D and tax credits for R&D, as shown in Table 1 of the Working Paper.

Revenue

Data Source Definitions and Limitations

2. OECD System of National Accounts (SNA) provides data on economic output by country, as a component of GDP, disaggregated by 2-digit ISIC Rev.4 industry code, including a code for the pharmaceutical industry.[1] The Supply-Use framework of the SNA also provides estimates of industry output, at ex-factory prices, by 2-digit ISIC Rev.4 code in supply and use tables (SUT).[2] Estimates are based on data reported to the OECD by national statistics agencies.

3. The IQIVIA data may result in an overestimation of net industry revenue because estimates may be based on list prices or average transaction prices and because, regardless of the price reported, there are no adjustments for off-invoice discounts or rebates, which can be confidential between sellers and buyers. Another limitation of IQVIA data is that country breakdowns may be blurred by cross-border sales, including but not limited to parallel exports/imports in EU Member States. The allocation of revenue to a country is driven by the geographic location of the purchaser (wholesaler, hospital or pharmacy) that supplies IQVIA with invoice data, not by the geographic location of the manufacturer that sells the product.

Methods

4. Industry revenue is aggregated from country-level to a global total and reported in two different views:

1. In a cross-section of the latest year available:
    a. In aggregate worldwide and broken down by country or geographic region in USD (2021 Q4 exchange rates); and,
    b. By disease area as a share in total revenue and the share of originators and generics in total revenue.
2. In a time-series for the preceding ten years:
    a. Also in aggregate worldwide and disaggregated by country or geographic region in USD at constant (2015) rates;

[1] Abbreviation for UN "International Standard Industrial Classification of all Economic Activities". The pharmaceutical industry is designated by ISIC Code 21 "Manufacture of basic pharmaceutical products and preparations." See United Nations (UN, 2003[12]).
[2] For SUT data, also see Discussion Paper DELSA/HEA(2019)14, presented at the joint Workshop of the Expert Group on Pharmaceuticals and Medical Devices and the Working Party on Health Statistics in October 2019. Data from SNA Supply-Use tables (SUT) have also been evaluated as a possible source for estimating total pharmaceutical expenditure.

b. By disease area as a share in total revenue and the share of originators and generics in total revenue; and,

c. By the top 15 companies with data available, by total sales revenue.

5. In OECD SNA data, output is valued at basic prices, excluding product taxes, trade costs, and the costs of delivery of the goods from producers to buyers. Although not identical to revenue, output is a concept that can be assumed to approximate revenue in the case of the pharmaceutical industry. It generally comprises goods sold and supplied for free; supplied to other establishments belonging to the same enterprise for use as intermediate inputs; goods retained by owners for own final consumption or own gross fixed capital formation; or goods used for payments in kind. Output is typically estimated using revenue less sales taxes, plus subsidies and change in inventory of finished and semi-finished goods.[3]

6. In IQVIA data, sub-categories of product types were available, and include: Innovative Branded Products, Unbranded Products, Non-Original Branded Products, Innovation Insights Not Assigned, Other Products, Non Rx Bound Products, and Vaccines. For the originator and generic aggregations, 'Innovative Branded Products' and 'Vaccines' were categorised as originators, and the remaining categories were classified as generics.

Net cash flows from financing activities

Data Source Definitions and Limitations

7. Data were extracted for the period 2005 to 2020, for which the OECD Capital Market Series dataset, Refinitiv Datastream contains financial statements for 50 087 unique non-financial firms listed on public stock markets, resulting in a total of 471 188 firm-year observations.

8. The sample covers 40 countries, including all OECD countries except Chile, Costa Rica, Poland, and Türkiye, as well as Brazil, China, India, Indonesia, Romania, and Russia.

Methods

9. The median and quartiles are computed to show the distribution across firms.

10. The pharmaceutical industry is defined as all firms assigned to NACE Codes 2110, 2120, and 7211. Pharmaceuticals are compared to four other high-tech industries and other sectors of the economy as described in Section 2 in the Working Paper. Financial and investment firms, such as firms that conduct trust, fiduciary and custody activities, asset management firms, and investment funds, are excluded from the sample.

11. The indicator is reported separately for the entire sample, OECD and non-OECD countries. Firms are assigned to countries based on the location of their headquarters. For firms listed on more than one stock exchange, only the primary listing is retained. The main criteria for including countries in the sample is the quality of data in the OECD-Orbis dataset used for other indicators based on firm-level data. To ensure consistency, the same countries are included in the Refinitiv Datastream sample and the OECD-Orbis sample. As the data providers for OECD-Orbis, provided by Bureau van Dyke (BvD) can change over time, so can the coverage and quality of the data. This is reflected by significant fluctuations in the number of observations over time. Accordingly, countries that have a more balanced sample are included in the analysis to avoid introducing additional biases. For this reason, OECD countries such as Poland and Türkiye are excluded from the sample. Countries such as Argentina, Chile, Costa Rica, Saudi Arabia, and South Africa are also excluded because sample sizes are too small.

[3] For definitions and non-technical discussions, see OECD Glossary of Statistical Terms (https://stats.oecd.org/glossary/detail.asp?ID=1968) and United Nations (UN, 2003[12]).

12. Aggregates are also analysed in sub-groups by firm size and firm age, where firm size is measured by total assets and firm age is defined as the difference between the current year and the year of incorporation as provided in the datasets. On both measures, firms are divided into four quartiles in each year, with the first quartile containing the smallest and youngest 25% of firms and the fourth quartile the largest and oldest 25% of firms.

13. To avoid introducing a survival bias towards successful companies, an unbalanced panel dataset is constructed. This means that, for the selected industries and countries, all available observations remain in the sample in each year and the composition of the sample can change over time. Such changes can reflect genuine entry and exit, i.e. the creation of new firms or failure of existing firms, but can also reflect changes in the sample coverage. This approach is preferred over using a balanced panel, i.e. retaining in the sample only firms with a complete set of consecutive annual observations, because it may be common for firms in the pharmaceutical industry to fail in early phases of their life cycles. Retaining only firms with a complete set of consecutive observations may bias estimates towards successful firms, because firms that fail during the period analysed would be excluded.

Direct subsidies for R&D

Data Source Definitions and Limitations

14. The OECD RDS provide aggregate estimates of R&D expenditure in the business enterprise sector (BERD) broken down by industry using ISIC Rev.4, which includes a category for the pharmaceutical industry,[4] and by source of funds. The disaggregation by source of funds identifies the following five distinct funding sources: the business enterprise sector (i.e. own funds of firms in the respective industry); the three domestic sectors: government, higher education, and private non-profit; and the rest of the world, which aggregates all funds received from foreign entities, including foreign governments and non-governmental funding agencies based abroad. This allows for identifying the portion of R&D that is funded by sources external to the industry.

15. OECD RDS provides a wide range of data on the resources devoted to R&D in all OECD countries and selected non-member economies.[5] These statistics are compiled from various sources, including national surveys by statistical offices in OECD countries. While national surveys may elicit microdata at the firm level,[6] they are typically designed to generate, in combination with census data, accurate estimates at higher levels of aggregation (e.g. at the level of industries or the entire economy of a country) and apply sampling and weighting methods accordingly. Microdata are not available for analysis.

Methods

16. Direct subsidies for R&D are defined as the difference between total BERD and BERD funded by the firms themselves, which is equal to the sum of BERD funded by domestic government, higher education, and private non-profit entities and by foreign entities. Direct subsidies are reported in two different views:

1. In a cross section of the latest year available, in aggregate for the countries for which data are available and broken down by country, in USD at current purchasing power parities (PPP). The

[4] See Note 6.

[5] See https://www.oecd.org/sti/inno/researchanddevelopmentstatisticsrds.htm

[6] Including, for example, basic demographic information (employment, industry of main activity, sales and type of ownership) together with detailed information about the firm's R&D, including information about R&D performed (intramurally) and funded (performed extramurally), the type of R&D performed (basic research, applied research, experimental development), sources of funding (e.g. own, other business, government), and the structure of R&D costs (e.g. labour, current consumption of goods services, capital) and R&D employment (OECD, 2020[11]) (OECD, 2020[11]) (OECD, 2021[2]).

cross section also compares the pharmaceutical industry to other research-intensive industries for which data are available.

2. In a time series for the past ten years, also in aggregate and broken down by country or geographic region, in USD at constant PPP.

17. In addition to the absolute amounts in USD at PPP, both views will also report direct subsidies as a percentage of total BERD.

Tax Credits for R&D

Methods

18. Tax credits are reported in terms of Government Tax Relief for R&D Expenditure (GTARD), i.e. forgone and refunded government tax revenue as a result of R&D tax credits (OECD, 2020[1]; OECD, 2021[2]), in the latest year in aggregate for the OECD countries for which data are available, and broken down by country. In this cross-section, the pharmaceutical industry is compared to other research-intensive industries for which data are available. The indicator is reported in two views:

- In absolute terms, in USD at current purchasing power parities (PPP); and,
- As a percentage of BERD.

19. No data are currently available for a longitudinal view or to estimate the implied marginal tax subsidy rates for R&D at the level of the pharmaceutical industry.

Activity

Profitability

Data Source Definitions and Limitations

20. Prior evaluations of the representativeness of Orbis found that the dataset provided good coverage of manufacturing industries, and the pharmaceutical industry more specifically, and that it is well suited for analyses that take a global perspective rather than comparing countries (Bajgar et al., 2020[3]; Kalemli-Ozcan et al., 2015[4]). However, country coverage is variable and tends to be better for European countries (Bajgar et al., 2020[3]). It should be noted that Orbis contains a relatively small number of non-listed pharmaceutical firms from the United States (see Annex B).

21. Data are extracted for the years 2005 to 2020 from Bloomberg, FactSet and Refinitiv Datastream, and for the years 2007 to 2019 from OECD Orbis. Data from Orbis for the years 2005 and 2006 are not used because of changes in sample composition that could bias trends. Orbis data for 2020 were not yet available at the time of analysis.

Methods

22. The profitability measures are constructed as described in Table A.1. All measures are reported in aggregate for the pharmaceutical industry for each year. The median and quartiles are computed to show the distribution across firms.

Table A A.1. Computation of profitability measures

Measure	Computation
Gross operating margin	(sales - cost of goods sold) / sales
Net operating margin	Earnings before interest and taxes (EBIT) / sales
ROA	Net income / total assets
Net ROA	ROA – COK, where cost of capital is the weighted average cost of equity and debt (WACC)

Note: At the firm-level, ratios are only computed for firms with non-zero denominators in a given year; for aggregates, values for all firms, including zeros, are aggregated before computing ratios.
Source: Authors

23. To assess a possible effect of the prevailing accounting treatment of R&D expenditure, a simulation is performed to capitalise R&D expenditure and amortise R&D assets, as described in Box A.1.

24. The profitability measures are reported separately for the entire sample, OECD and non-OECD countries. Aggregates are also analysed in sub-groups by firm size and firm age, where firm size is measured by total assets and firm age is defined as the difference between the current year and the year of incorporation as provided in the datasets. On both measures, firms are divided into four quartiles in each year, with the first quartile containing the smallest and youngest 25% of firms and the fourth quartile the largest and oldest 25% of firms.

25. The same unbalanced panel datasets are constructed with an identical geographic scope as for all other financial indicators based on firm-level data, as described above in the Section titled “Net cash flows from financing activities.” Across both datasets, the pharmaceutical industry is defined as all firms assigned to NACE Codes 2110, 2120 and 7210. Pharmaceuticals are compared to four other high-tech industries and other sectors of the economy as described in Section 2 of the Working Paper.

26. The OECD Capital Market Series dataset, Refinitiv Datastream (referred often as ‘Refinitiv Datastream’) and the OECD-Orbis Corporate Finance dataset (referred often as ‘OECD Orbis’) datasets were cleaned and prepared for analysis according to the steps described below.

Box A A.1. Capitalising R&D costs in profitability estimates

Accounting treatment of R&D expenditure in the pharmaceutical industry

With some exceptions, both IFRS and US GAAP accounting standards require that R&D costs in the pharmaceutical industry be recognised as expenditure in the period in which they were incurred (see Box A A.2). However, to the extent that R&D is successful and creates revenue-generating assets, this accounting treatment can lead to an understatement of a firm's asset balance. It has been argued that this issue could result in misleading profitability measures for the pharmaceutical industry, in particular an overstatement of return on assets because of a lower asset balance (see, for example, (Damodaran, n.d.[5])). Conversely however, failure to capitalise R&D expenditure that contributes to creation of future assets also implies an overstatement of current expenditure, and therefore an understatement of operating margins and a reduced tax burden on income. The net effect on profitability in any given period depends on which of the two opposing effects is stronger.

Simulating return on assets with capitalised R&D expenditure

To test whether accounting conventions may indeed bias profitability measures, a simulation was performed in which 100% of R&D expenditure was capitalised in the period incurred and the resulting asset balances amortised linearly over ten years. That is, any R&D expenditure reported in a given year is reduced to zero in the profit and loss statement and added to assets on the balance sheet. In each of the ten subsequent years, one-tenth of the additional asset balance is recognised as amortised expenditure. Operating profits are then adjusted for the difference between these two items and any incremental profits reduced by 25% to account for additional taxation of corporate profits. Adjusted net margins and asset balances are then used to recalculate returns on assets. Although patent protection and other forms of exclusivity have been estimated to effectively shield a new medicine from generic competition for approximately 12-14 years from marketing authorisation (Copenhagen Economics, 2018[6]), this simplified methodology yields a conservative estimate of the possible impact of accounting conventions on profitability. This is because only a small proportion of R&D projects result in an authorised medicine. For example, over the period 2000 to 2015 the probability of successful marketing authorisation was estimated to range, on average, from 14% for medicines entering phase 1 to 59% for those in phase 3, with significant variation across therapeutic areas (Wong, Siah and Lo, 2018[7]). As a result, only a portion, rather than 100% of R&D expenditure, would meet the requirements for capitalisation. In addition, R&D projects can fail at any time prior to marketing authorisation, which implies that asset balances resulting from earlier expenditure on projects that failed at a late development stage would likely have to be written off long before a time horizon of ten years is reached.

Source: Authors based on sources cited in the text.

OECD Capital Market Series dataset, Refinitiv Datastream

27. Data are extracted for the years 2005 to 2020 for all publicly listed firms. Financial and investment firms, such as firms that conduct trust, fiduciary and custody activities, asset management firms, and investment funds, are excluded from the sample. Only one observation for each unique firm-year combination is retained. When firms are listed on more than one stock exchanges, only the primary listing is included.

OECD-ORBIS Corporate Finance dataset

28. Data are extracted for the years 2005 to 2019 for all unlisted firms only. Preparation and cleaning of the dataset comprises the following steps to retaining only one observation for each unique firm-year combination.

1. Using only annual financial statements by:
    a. Using only financial statements covering a 12-month period or, if not available, those that cover 11 months and annualising flow variables (i.e. items from the profit and loss and cash flow statements).
    b. When multiple observations for the same year and firm exist, using the financial statement with the closing date closest to the end of the corresponding calendar year.
    c. Assigning closing dates that fall between 1 January and 30 June of a year to the previous calendar year.
2. Selecting accounting standards by giving preference to financial statements filed according to the International Financial Reporting Standards (IFRS) over those filed according to US Generally Accepted Accounting Principles (US GAAP), and preference to US GAAP over other local and unknown accounting practices. All observations with unspecified accounting standards are dropped.
3. Giving preference to published annual reports are preferred to local registry filings. All observations with unspecified filing types are dropped.
4. Retaining only unconsolidated financial statements to avoid double counting of firms that consist over several legal entities by:
    a. Selecting the appropriate consolidation codes that denote unconsolidated accounts.
    b. Using the Orbis ownership file to identify direct controlling and ultimate shareholders of all firms and excluding firms with only consolidated accounts that are direct controlling shareholders or ultimate shareholders of other firms.

29. In addition, firm-year observations are dropped if total assets or tangible fixed assets are reported to be negative; and during the entire period, the number of employees of the firm remains below five.

Financial Data Disaggregation

30. To help elucidate the heterogeneity found in the firms that make up the pharmaceutical industry, the following disaggregation exercise was carried out. A geographically-stratified sample of 1,000 firms was created using the list of pharmaceutical firms from OECD Orbis and Refinitiv Datastream data (see Annex B). Each firm was then manually assigned to a company category, based on the publicly available information on their business operations. The categories used were: *contract research organisations (CRO), Large Pharmaceutical Companies (top 15 firms by sales revenue), Other Research-Based Pharmaceutical Companies, and Manufacturers (including Generics Manufacturers).* A number of indicators were then computed for the firms in the sample, disaggregated according to the categories outlined above and thereby allowing for comparison across company type.

Research & development (R&D) expenditure

Data Source Definitions and Limitations

31. The OECD Research & Development Statistics (RDS) are compiled from various sources, including national surveys by statistical offices in OECD countries. While national surveys may elicit microdata at

the firm level,[7] they are typically designed to generate, in combination with census data, accurate estimates at higher levels of aggregation (e.g. at the level of industries or the entire economy of a country) and apply sampling and weighting methods accordingly. Microdata are not available for analysis.

32. As discussed in Section 3.2.5 of the Working Paper, expenditure can be assigned to a given industry based either on the "main activity" of each of the firms in an industry or on a "industry orientation"-basis. The distinction between the two approaches is relevant when there are firms in an economy that undertake R&D for a diversified set of products, or when firms that fund and perform R&D are not the same. In pharmaceutical R&D this may occur, for example, when contract research organisations (CROs) are commissioned by pharmaceutical firms to conduct clinical trials. In this case the R&D expenditure incurred by CROs may be classified in ISIC Code 72 "Scientific research and development" by "main activity" based on CROs' main activity being assigned to this category, while the "industry orientation" of the R&D by CROs may fall into Code 21 "Manufacture of pharmaceuticals, medicinal chemical and botanical products." R&D expenditure in the pharmaceutical industry classified by "main activity" may therefore be understated because it does not include pharmaceutical R&D performed by research firms and, at the same time, overstated because it may include R&D expenditure by pharmaceutical firms that do not contribute directly to the development of medicines.

33. When using the OECD Orbis and Refinitiv Datastream data for R&D expenditure, there are some limitations with using accounting standards. Geographic discrepancies may occur for instance, when a multinational firm has a subsidiary that conducts R&D in one country but files its consolidated financial statements in another country, or when a firm acquires an R&D project from another country. Temporal discrepancies result from accrual-based accounting. Although under both, IFRS and US GAAP, R&D expenditure is generally recognised as an expense in the year it is incurred, firms may also capitalise expenditure and recognise intangible assets on their balance sheets if the underlying R&D project meets a number of conditions (see Box A A.2). R&D expenditure in a given year may thus exclude R&D activity that took place in that year but was capitalised as part of an intangible asset; it may, on the other hand, include asset impairment charges or write-downs related to existing intangible assets, while the R&D activity that created the asset took place and was capitalised in prior years. These differences may be material, particularly when firms acquire R&D projects that are in progress and recognise these assets at acquisition cost on their balance sheets, or when they are subsequently required to reduce the book value of such assets.

[7] Including, for example, basic demographic information (employment, industry of main activity, sales and type of ownership) together with detailed information about the firm's R&D, including information about R&D performed (intramurally) and funded (performed extramurally), the type of R&D performed (basic research, applied research, experimental development), sources of funding (e.g. own, other business, government), and the structure of R&D costs (e.g. labour, current consumption of goods services, capital) and R&D employment (OECD, 2020[1]).

Box A A.2. Accounting treatment of R&D expenditure

Both of the two major accounting standards, IFRS and US GAAP, generally require that internal R&D expenditure of pharmaceutical firms be recognised as an expense in the current period. This is mainly related to the uncertainty around the ultimate outcomes of R&D at the time when it occurs, i.e. a new medicine in development may or may not be authorised for marketing, and it is therefore uncertain whether the project will result in a revenue-generating asset. However, there are exceptions to this rule and some differences between the standards. In particular, IFRS provides broad rules about when firms must capitalise R&D expenditure. These state that R&D costs should be recognised as intangible assets if a firm can demonstrate all of the following (see IFRS Foundation (IFRS, 2021[8])):

- the technical feasibility of completing the intangible asset so that it will be available for use or sale;
- its intention to complete the intangible asset and use or sell it;
- its ability to use or sell the intangible asset;
- how the intangible asset will generate probable future economic benefits. Among other things, the entity can demonstrate the existence of a market for the output of the intangible asset or the intangible asset itself or, if it is to be used internally, the usefulness of the intangible asset;
- the availability of adequate technical, financial and other resources to complete the development and to use or sell the intangible asset; and,
- its ability to measure reliably the expenditure attributable to the intangible asset during its development.

In addition, when firms acquire in-progress R&D projects, these are generally recognised as assets at acquisition cost on their balance sheets. Such balances may include significant goodwill beyond accumulated R&D costs.

Source: Authors based on IFRS Foundation (IFRS, 2021[8]) and KPMG Advisory (KPMG, 2021[9])

Methods

34. Using aggregate data from OECD R&D Statistics, this indicator replicates the computations of aggregate industry R&D expenditure (BERD) and R&D intensity published already in Health at a Glance (OECD, 2021[10]). In addition to the cross-sectional view shown in Health at a Glance, it will show trends over time. The indicator is reported in terms of:

a. BERD in the pharmaceutical industry in absolute terms from 2010 to 2019 or the latest year available, in USD at constant purchasing power parities (PPP), in aggregate for all countries with data and individually for the top countries;

b. R&D intensity, i.e. BERD as a share of gross valued added (GVA) in the pharmaceutical industry and comparison industries (see Section 2 of the Working Paper), in 2019 or the latest year available, for each country with data and as an OECD aggregate across all OECD countries with data.

35. Using firm-level data, the indicator is reported in terms of R&D intensity in the pharmaceutical industry only, defined as R&D expenditure as a share of revenue, for the years 2005 to 2020. Refinitiv Datastream and OECD-Orbis do not report GVA consistently so that it is not possible to use GVA as denominator in a computation of R&D intensity. The Refinitiv Datastream and OECD-Orbis samples are cleaned and prepared for analysis as described in Section 3.2.2 of the Working Paper. The geographic

scope covers a total of 40 countries, including all OECD countries except Chile, Costa Rica, Poland, and Türkiye, as well as Brazil, China, India, Indonesia, Romania, and Russia. The indicator is reported separately for OECD and non-OECD countries. Because there is a large number of missing observations in the variables that contain R&D expenditure in firm-level data, especially in Orbis, a missing value in R&D expenditure for a firm that does report other key financial variables in the same year is assumed to represent zero R&D expenditure. Although this may have the effect of understating R&D intensity somewhat, retaining only observations with non-missing R&D expenditure would likely have a stronger opposite effect. As a sensitivity analysis, R&D intensity is also estimated for firm/year pairs that have non-missing R&D expenditure values only.

Number of products in development

Data Source Definitions and Limitations

36. Springer Nature provided the OECD with historical 'snapshots' of the AdisInsight database as at 31 December of each year for the years 2011 to 2020. A current snapshot is provided with the status on 31 December 2021, and an update will be provided for 31 December 2022.

37. For the years 2011 to 2020, the first five variables in Table A.2 were used. The current datasets available for the years 2021 and 2022 are more detailed and four additional variables are used for analysis (also see Table A A.2).

Table A A.2. AdisInsight variables used for analysis

Variable	Description	2011-20 datasets	2021 dataset
Adis Number	Unique product ID	√	√
Country of development	Country/countries in which the compound is being developed, i.e. country/countries that host lab studies or clinical trials.	√	√
Indication	Disease(s) for which the compound is being developed.	√	√
Phase of development	Highest development phase in a given year (from pre-clinical to pre-market authorisation).	√	√
Development status	Identifies products in active development or with other statuses (discontinued, withdrawn, suspended, unknown).	√	√
Development rank	Numeric rank based on development status and phase; ranks 1 and 2 denote products that have received marketed authorisation, ranks 2 to 14 products that are in active clinical or pre-clinical development.		√
Orphan designation	Binary variable with value 1 if an indication for which the compound is in development is an orphan disease.		√
New Molecular Entity (NME) designation	Binary variable with value 1 if a compound is an active substance that has not been previously authorised in the European Union or the United States.		√

Variable	Description	2011-20 datasets	2021 dataset
Involvement Role	Entities involved in product development and their respective role (e.g. originator, current IP owner, licensee, etc.).		√

Source: Authors based on Springer Nature

Methods

38. The number of unique combinations of product IDs and indications in AdisInsight were counted to compute the number of product development projects. The number of product profiles, i.e. the number of unique product IDs, was also counted. The indicator focuses on active development projects, i.e. products that are in a preclinical or clinical development phase for a given indication and whose development status is reported to be “active” on the last day of the year.

39. The number of active development projects in the latest year was broken down in a manner similar to that of the WHO Global Observatory on Health R&D. A geographic breakdown was also provided, so that the number of active development projects is broken down by:

- Disease area to which the indication belongs;
- Development phase; and,
- Country of development

40. In addition, the proportion of product development projects that originated in the pharmaceutical industry was computed within each category of the breakdowns above. Projects were classified as originating in the industry if the originator is a firm that is part of the pharmaceutical or biotechnology industry. AdisInsight identifies all organisations that contribute to a development project, the organisation type and their respective roles. Organisation types include biopharmaceutical firms, biotechnology firms, hospitals, and universities, which are aggregated into broad industry and non-industry categories. Roles include originator, owner, and various types of licensees. An “originator“ is defined as the organisation that initiates the product or development project and is usually the initial owner of the intellectual property rights (IPRs); this organisation is never updated in a product profile and there is no geographic location associated with the role. An “owner“ is defined as the organisation that currently owns the IPRs in a product; the owner may be the same as the originator but ownership can change over time as IPRs are transferred or acquired, or if the entire originator organisation is acquired by another entity.

41. For the years 2011 to 2020, two longitudinal analyses are provided. First, the breakdown of the number of active development projects by disease area was plotted over time. Second, for each year, the total number of development projects was broken down into the following categories:

- New (for the years 2012 to 2021): the product-indication combination was not in active development in the prior year;
- Active in the same phase (2011 to 2020): the product-indication combination was still in active development in the same phase in the subsequent year;
- Active progressed (2011 to 2020): the product-indication combination was still in active development when a subsequent phase or development was completed (i.e. the product received marketing authorisation in at least one country) in the subsequent year;
- Newly authorised (2011 to 2020): the product-indication combination was still in active development and listed as authorised / launched in the subsequent year; and,
- Discontinued (2011 to 2020): the product-indication combination was neither in active development nor completed in the subsequent year.

42. The datasets were cleaned and prepared for analysis as follows:

- Products were dropped from the dataset if:
  - they were classified as "diagnostic agents", so that only medicines and vaccines are retained;
  - the Adis Number (unique ID) was missing; or,
  - data in all fields other than the Adis number were missing.
- Entries from the country variable are dropped if they refer to a region and cannot be unambiguously assigned to a single country.
- The following variables were created for all datasets:
  - An "OECD" flag indicating whether a given country is an OECD country.
  - An "Industry" flag that identifies the following organisation types, as defined by Springer Nature, that can be involved in product development and have a profit motive, as part of the pharmaceutical industry: technology transfer, agency, biopharmaceutical, CRO, company, biotechnology, investor, manufacturer, pharmaceutical, technology provider, distributor, diagnostic, nutraceutical, medical device.
  - Two "Industry" flags associated with each product: one that flags the product if all originators belong to the pharmaceutical industry and one that flags products with at least one originator in the pharmaceutical industry.
  - Two "Industry" flags associated with each product based on the current IP owners of a product.
  - The indication in the "Indication" variable is mapped to higher-level disease categories at two hierarchical levels (e.g. infectious vs. non-communicable diseases at level 1; broad disease categories such as malignant neoplasms, cardiovascular diseases, or neurological conditions, etc. at level 2) using the automated data mining tool for the standard classification of health categories developed by the WHO Global Observatory on Health R&D[8]; where the data mining tool did not provide a suggestion, the Secretariat mapped the indication manually.
  - An "Orphan Designation" flag identifying the drug profiles associated with an orphan disease (i.e. where the "orphan indication" column was not empty).
  - For the development phase indicator, only the highest available phase for each combination of product ID X indication was retained; a given product can be simultaneously in different development phases in different countries.
- For the historical datasets only, the following variables were created:
  - A development rank variable identical to the 2021 dataset (see Table A A.2), based on the combination of development status and phase.
  - A numerical variable defined as the difference in development rank between year y and y-1, to categorise products according to their progression between development statuses and phases from one to the subsequent year.

Number of clinical trials and people enrolled in clinical trials

43. Seventeen data providers currently feed registration data from primary registers into ICTRP, including from the Australian New Zealand Clinical Trials Registry (ANZCTR); the Brazilian Clinical Trials Registry (ReBec); the Chinese Clinical Trial Register (ChiCTR); ClinicalTrials.gov (United States); Clinical Trials Registry - India (CTRI); EU Clinical Trials Register (EU-CTR); Japan Primary Registries Network (JPRN).[9]

[8] See https://www.who.int/observatories/global-observatory-on-health-research-and-development/classifications/health-categories.
[9] For the full list of primary registers and data providers, see https://www.who.int/clinical-trials-registry-platform/network/data-providers.

44. The ICTRP data fields listed in Table A.3. are analysed to generate this indicator. ICTRP contains no coded data about the conditions or disease area studied in a given trial, the product or intervention studied, the country in which the trial is conducted, or about the entity that finances it. However, various free-text fields contain unstructured information about the intervention studied and trial sponsorship.

Table A A.3. ICTRP data fields analysed to generate the number of clinical trials

Field labels	Field codes	Definitions	Purpose in OECD analysis
Countries of Recruitment	country2	The countries from which participants will be, are intended to be, or have been recruited.	Assign trials to countries in which they take place (done by WHO R&D Observatory).
Date of enrolment	date_enrolment	Anticipated or actual date of enrolment of the first participant.	Identify trial start year.
Health condition(s) or problem(s) studied	hc_freetext hc_code hc_keyword	Primary health condition(s) or problem(s) studied (e.g., depression, breast cancer, medication error). If the study is conducted in healthy human volunteers belonging to the target population of the intervention (e.g. preventive or screening interventions), this field should contain the particular health condition(s) or problem(s) being prevented. If the study is conducted in healthy human volunteers not belonging to the target population (e.g., a preliminary safety study), the field should contain an appropriate keyword.	Assign trials to a disease area (done by WHO R&D Observatory).
Intervention(s)	i_freetext i_code i_keyword	Specific name of the intervention(s) and the comparator/control(s) being studied, using the International Non-Proprietary Name (INN) if possible (not brand/trade names). For an unregistered drug, the field can contain the generic name, chemical name, or company serial number. If the intervention consists of several separate treatments, the filed should contain a list of all of them separated by commas (e.g., "low-fat diet, exercise").	Distinguish between trials of medicines and trials of other products or interventions.
Primary sponsor	primary_sponsor	The individual, organisation, group or other legal entity, which takes responsibility for initiating, managing and/or financing a study.	Identify the primary sponsor of a given trial and distinguish between industry-sponsored trials and other trials.
Secondary sponsor	sponsor_name	Additional individuals, organisations or other legal persons, if any, that have agreed with the primary sponsor to take on responsibilities of sponsorship. A secondary sponsor may have agreed: - To take on all the responsibilities of sponsorship jointly with the primary sponsor; or - To form a group with the primary sponsor in which the responsibilities of sponsorship are allocated among the members of the group; or - To act as the sponsor's legal representative in relation to some or all of the trial sites; or - To take responsibility for the accuracy of trial registration information submitted.	Identify secondary sponsors of a given trial and distinguish between trials with and without industry involvement.
Source(s) of monetary or material support	source_name	Major source(s) of monetary or material support for the trial (e.g., name of funding agency, foundation, firm, etc.). Note: This field is frequently empty or contains a reference to primary or secondary sponsor(s).	Identify secondary sponsors of a given trial and distinguish between trials with and without industry involvement.
Public title	public_title	Title intended for the lay public in easily understood language.	Distinguish between trials of medicines and trials of other products or interventions.
Recruitment status	recruitment_status	Recruitment status of this trial. The field should contain one of the following values: - Pending: recruitment of participants has not yet commenced - Recruiting: open to recruitment	Distinguish active trials from trials that have not yet started or have been completed.

		- Suspended: recruitment has been temporarily stopped - Complete: participants are no longer being recruited. The trial is closed. Participants may, however, still be in follow-up. (Note: some registries may choose to divide this into 2 options: Complete: follow-up continuing; Complete: follow-up complete.) - Other	
Scientific title	scientific_title	Scientific title of the study as it appears in the protocol submitted for funding and ethical review. Include trial acronym if available.	Distinguish between trials of medicines and trials of other products or interventions.
Secondary sponsors	sponsor_name	Additional individuals, organizations or other legal persons, if any, that have agreed with the primary sponsor to take on responsibilities of sponsorship. A secondary sponsor may have agreed: - To take on all the responsibilities of sponsorship jointly with the primary sponsor; or - To form a group with the primary sponsor in which the responsibilities of sponsorship are allocated among the members of the group; or - To act as the sponsor's legal representative in relation to some or all of the trial sites; or - To take responsibility for the accuracy of trial registration information submitted.	Identify trials whose primary sponsors are entities that are not part of the industry but to which industry contributes in a secondary role.

Note: The field definitions are based on the guidelines published by WHO; actual values contained by these fields may or may not comply with these guidelines.
Source: Authors based on ICTRP Data Codebook, Version 1.3.1

45. The WHO Observatory of Global Health R&D analyses free-text fields to generated coded variables for diseases, trial registration and recruitment dates and countries and provides the coded datasets to the OECD. The OECD Secretariat developed text-mining algorithms to identify trials of medicines and to identify industry-sponsored trials using free-text fields, discussed below.

Methods: Number of Clinical Trials

46. The number of trials, i.e. the number of unique trial IDs, in ICTRP were counted to compute the number of trials. The number is expressed in terms of trials that were actively recruiting participants or have been completed (referred to as "active" trials); it excludes trials that had been registered but for which recruitment had not begun. The initial time period analysed was 2010 to 2021; only trials registered in this time period were analysed.

47. The total number of new trials was computed per year by assigning each trial to a start year using the date of enrolment of the first participant. The number of active trials was identified by analysing the field "Recruitment status". Active trials were those that were open for recruitment or had been completed; the number of active trials thus represents the cumulative number of all trials that had begun in a given year and had since been completed or were still open for recruitment in the latest year. While WHO provides guidelines for coding of recruitment status in primary trial registers that feed data to ICTRP (see Table A A.3.), there is some variability across trial registries in how recruitment status is coded, and which precluded a more detailed analysis. In particular, the field does not distinguish consistently between trials in which recruitment has been completed but follow-up is ongoing from those trials in which follow-up is complete.

48. All trials were initially disaggregated into trials of medicines vs trials of other products or interventions, and into trials with a primary sponsor from industry, and trials with and without industry involvement (see below for definitions). Active trials of medicines with industry involvement were further broken into the same categories used by the WHO Global Observatory on Health R&D, by:

- Disease area or condition; and,
- Country or geographic region, breaking down the total into: OECD total; Europe, Japan, the United States, other OECD; and China, Brazil, India, other non-OECD.

49. The Secretariat developed text mining algorithms to identify trials of medicines, using the public and scientific titles and interventions fields, and to identify trials that were industry-sponsored or had industry involvement, using the text fields that identified the primary and secondary sponsors as well as the source of financial support. Further details about this algorithm are in Box A.3.

50. A trial of a 'medicine' was determined by if the intervention described in the *intervention* or *title* fields was identified by the text-mining algorithm as a 'medicine'. The definition of 'medicine' was adapted from the regulatory definition used by the European Medicines Agency (EMA), and excluded products used only for diagnosis:[10] *A substance or combination of substances that is developed or produced with an intention to treat or prevent a disease, or to restore, correct or modify physiological functions by exerting a pharmacological, immunological or metabolic action.* This includes biological medicines[11] and cell therapies in which the cells used have been modified; food or dietary supplements with an intended therapeutic effect; and therapeutic substances used in traditional Chinese medicine (TCM) or Ayurvedic medicine. It does not include convalescent plasma, blood transfusions, or cell transplants.

- "Industry-sponsored" if its primary sponsor was identified as a for-profit firm by the text mining algorithm.
- Having "industry involvement" if the primary sponsor, secondary sponsor, or the entity named as the funding source was identified as a for-profit firm by the text-mining algorithm.

[10] The EMA definition of "medicinal product" also includes substances used for diagnosis: "A substance or combination of substances that is intended to treat, prevent or diagnose a disease, or to restore, correct or modify physiological functions by exerting a pharmacological, immunological or metabolic action" (see https://www.ema.europa.eu/en/glossary/medicinal-product).

[11] A term generally used in pharmaceutical regulation to refer to broad category of medicines whose active ingredients are produced using living organisms.

Box A A.3. Identifying clinical trials of medicines and industry-sponsored trials

Text mining algorithms were developed to identify a) trials of medicines and b) industry-sponsored trials among all trials registered in ICTRP.

Trials of medicines

To identify trials of medicines, an algorithm analyses the fields *Intervention* and *Scientific Title*. Based on both fields, the algorithm generates a binary variable with the values "medicine" or "non-medicine".

Two lists of keywords were created manually. One contains words that frequently appear in descriptions of the intervention or in scientific titles of trials of medicines, and the other one words that are associated with trials of other types of interventions. In addition, the algorithm searches for a list of proper names of generic medicines and common endings of generic names. The algorithm follows the steps below to compare the words in these lists to all words that appear in the fields *Intervention* and *Scientific Title*, computing a score that represents a level of confidence that a given trial is a trial of a medicine.

1. The algorithm evaluates the *Intervention* field first, which is generally more indicative than the *Scientific Title*, because it specifically describes the medicines, devices or techniques investigated in the trial.
    a. If a word exists in the *Intervention* field that is identical to one of the words in the keyword lists or the list of proper names or name endings, it is added with a confidence score of 1 (the highest possible score) to a list of checked keywords.
    b. For every keyword included in the keyword lists that does not appear in the *Intervention* field, the algorithm computes the Levensthein Distance between the keyword and each word in the field to quantify the similarity of every word in the field to the keywords and assigns a score computed as 1 less one-tenth of the Levensthein Distance. For example, this will assign a score close to but less than 1 to the word "drugs" because the word "drug" is in the keyword list. For each keyword, it retains only the highest scoring word from among all words found in the intervention filed that scored at least 0.8. This word and its score is added to the list of previously checked keywords (generated in step 1.a). *Note: Step 1.b is only done for keywords, but not for generic names or endings.*
    c. Using the five highest scoring words in the list of all words with an individual score of at least 0.8, the algorithm computes an overall confidence score of the trial being a trial of a medicine by adding the scores of all keywords associated with medicines and subtracting the scores of all keywords not associated with medicines. A higher score thus indicates a higher confidence that a trial investigates a medicine. If the result of this computation is smaller than -0.1 the trial is classified as non-medicine trial; if it is greater than or equal to 0.8, it is classified as a medicine trial.
2. If the result of this computation based on the *Intervention* field is equal to or greater than -0.1 but lower than 0.8, the algorithm repeats the same process described in Step 1 for the *Scientific Title* field. Scores computed for the *Scientific Title* field are added to/subtracted from the scores previously computed for the *Intervention Field*.

After both steps, a trial is classified as a medicine trial only if the final score is greater than or equal to 0.8. All final scores below 0.8 remain in a default non-medicine category.

Industry-sponsored trials

The identification of industry-sponsored trials follows the same logic described above for trials of medicines. An algorithm evaluates independently the fields *Primary Sponsor*, *Secondary Sponsors* and *Source Support*.

For each of these fields, the algorithm generates a binary variable with the value "industry" or "non-industry" based on a list of keywords that are associated with for-profit firms and keywords associated with other types of entities, such as universities, research institutions, hospitals, and governmental bodies. Keywords associated with firms are added to the score while keywords not associated with firms are subtracted from the score.

A trial is classified as "non-industry" if the overall score computed for the field is less than 0. All scores above or equal to 0 are classified as "industry".

Accuracy of results

These two algorithms were validated with two random samples of 500 trials each drawn from the entire ICTRP dataset. All trials in these two samples were coded manually by the Secretariat as medicine/non-medicine and industry-sponsored/non-industry. The algorithms generally achieve an accuracy of above 90%, except in non-industry-sponsored trials as shown below:

	Medicine trials		Industry-sponsored trials	
	Sensitivity	Specificity	Sensitivity	Specificity
Sample 1 (n=500)	91.6%	90.8%	100.0%	85.0%
Sample 2 (n=500)	90.5%	93.3%	100.0%	86.9%

In identifying industry-sponsored trials, the main type of error is a false-positive (i.e. low specificity of approximately 85%). These errors stem from trial sponsor fields only containing proper names, either of natural persons who are investigators in a trial or of legal entities who take a sponsorship role but containing no keywords that indicate whether the name is associated with a firm or another type of entity. The algorithm classifies these as industry-sponsored although it is unclear whether that is the case.

In identifying medicine trials, sensitivity and specificity were between 91% and 93%, meaning that false positives and false negatives occur at similar rates in less than 10% of cases. Both types of errors are caused by *Intervention* fields that contain little or no information; are ambiguous; or describe not only the intervention that is investigated but also medicines, devices or procedures used across the intervention and control groups so that it is not immediately clear which of these is the intervention of interest. Although the algorithm also analyses the *Scientific Title* field, this field tends to be less informative and often only allows a trial to be associated with a medicine through identification of non-generic proper names, including experimental names, for which no controlled reference nomenclatures exist.

Source: Authors

Methods: Number of people enrolled in clinical trials

51. The median number of people enrolled in clinical trials is computed by observing the ICTRP variable trial sample size and obtaining the 50th percentile value. The sample is non-parametric with no resemblance to a normal distribution and notable skewness and kurtosis, which favoured calculation of the median as an output over the mean.

52. The indicator is disaggregated in the same way as the number of trials, as discussed above. That is, the median number of people enrolled in clinical trials is computed separately for trials of medicines and trials of other products or interventions; industry-sponsored trials and trials that are not sponsored by the industry; and, for industry-sponsored trials of medicines, by disease area or condition and country or geographic region.

53. The sample was further restricted through the identification of Interventional and Observational trials. The ‘study_type’ variable, defined as the type of study and design of the trial, was mapped to either Interventional or Observational Trials. Observational trials, which included large-scale post-market surveillance, retrospective cohort studies, etc. were removed to focus on pharmaceutical interventional trials with designated participants and a frame of observation. The variable ‘phase’ was cleaned and categorised into Phase 1 – 4 to help assess the number and size of trials at various stages of development. Additional cleaning was undertaken to remove trials at the lower and higher bounds (i.e. manual review of trials near 0 or above 50,000 participants).

54. The sample size is defined as the actual number of people enrolled in a given trial or, if the actual number is not reported, the number of subjects planned. While ICTRP contains the two separate data fields for the variable "sample size"[12], only a subset of trial records contain data about the actual number people enrolled and this information may become available in the dataset with a significant time lag after a trial is registered and conducted.

1. "Target_size": defined as the number of participants that this trial plans to enrol; and
2. "Results_actual_enrollment": defined as the number of participants that the trial has enrolled.

[12] See ICTRP Data Codebook, Version 1.3.1.

Annex B. Firm-level data

55. The tables below show the number of unlisted and listed pharmaceutical firms included in the OECD-Orbis Corporate Finance dataset and OECD Capital Market Series dataset, Refinitiv Datastream samples.

Table A B.1. Number of unlisted pharmaceutical companies in the OECD-Orbis Corporate Finance sample

Years 2005 to 2019

	'05	'06	'07	'08	'09	'10	'11	'12	'13	'14	'15	'16	'17	'18	'19
Australia	19	20	17	16	19	27	24	35	41	48	54	63	60	58	39
Austria	118	124	139	140	142	148	156	164	163	167	168	169	171	172	85
Belgium	228	244	262	281	288	303	319	345	360	370	387	395	421	431	373
Brazil	17	30	29	26	26	26	26	47	49	58	72	86	83	64	35
Canada	11	11	13	13	9	4	5	7	9	12	9	9	11	12	5
China	3247	4524	4950	5285	4980	4092	4116	4072	5187	6835	1513	1630	1582	1377	239
Colombia	326	469	479	521	491	518	577	596	781	839	741	732	868	963	1134
Czech	54	61	66	72	78	76	79	87	86	85	79	80	77	74	32
Denmark	91	102	111	119	127	131	137	143	142	142	142	157	175	187	198
Estonia	16	21	26	30	36	40	42	45	52	54	58	60	62	57	57
Finland	36	34	45	49	51	58	59	57	61	61	62	66	69	74	75
France	473	472	461	488	529	541	602	630	620	595	503	492	502	474	322
Germany	537	774	787	812	796	794	822	737	702	698	705	711	739	726	265
Greece	61	66	72	72	74	79	82	83	83	89	86	91	96	101	42
Hungary	192	63	127	131	215	203	217	238	254	270	281	289	294	296	265
Iceland	18	19	20	20	20	24	22	22	24	24	24	26	27	28	29
India	246	259	263	275	280	275	367	689	1738	1678	1684	1576	1380	983	102
Indonesia		2	3	3	2	1	2	2	1	1	2	1	1	1	1
Ireland	121	131	138	142	141	140	138	150	154	152	153	163	163	165	102
Israel	1	3			1	1	1	12	11	4	3	6	5	2	3
Italy	435	450	559	592	611	629	643	642	659	667	709	750	779	818	744
Japan	95	98	103	114	115	124	145	152	153	155	159	168	175	158	124
Korea	306	320	316	318	287	321	351	382	388	431	459	494	504	470	391
Latvia	15	15	18	19	21	23	25	28	31	32	34	37	38	45	42
Lithuania	3	3	5	6	7	7	7	7	8	10	11	11	11	11	8
Luxembourg	5	7	7	5	13	15	10	10	14	12	13	17	22	22	12
Mexico	8	10	6	6	4	11	7	14	15	16	22	33	37	25	12
Netherlands	184	187	192	190	202	211	233	270	309	336	368	402	431	498	408
New Zealand	1	1	2	4	4	7	9	20	30	26	15	12	12	13	13
Norway	51	56	62	61	68	73	69	71	71	75	77	82	79	86	98
Portugal	113	120	119	129	134	137	141	152	156	154	161	169	170	168	163
Romania	123	109	121	111	103	106	109	118	127	130	133	136	131	131	130
Russia	1269	1303	1177	1037	972	981	970	1109	1092	1024	1059	1079	1098	1124	1070
Slovak Republic	13	13	12	14	18	21	22	25	23	23	22	21	22	20	19
Slovenia	5	6	6	8	9	51	51	50	53	49	45	41	46	44	49
Spain	383	412	388	448	461	469	477	483	494	498	509	527	534	522	345
Sweden	119	99	86	84	88	101	101	109	113	117	112	117	115	126	138
Switzerland	6	4	2	2	2	2	2	3	4	4	8	7	6	4	2
United Kingdom	937	1013	1066	1174	1301	1439	1635	1789	1912	2027	2146	2258	2321	2413	1403
United States	104	95	74	79	89	99	157	193	204	136	106	139	160	139	79

Source: OECD analysis based on OECD-Orbis database

Table A B.2. Number of publicly listed pharmaceutical companies in the OECD Capital Market Series dataset, Refinitiv Datastream sample

Years 2005 to 2020

	'05	'06	'07	'08	'09	'10	'11	'12	'13	'14	'15	'16	'17	'18	'19	'20
Australia	65	71	68	70	67	64	64	64	65	71	72	75	78	81	79	85
Austria													1		1	1
Belgium	5	7	8	7	8	7	6	5	6	6	8	9	9	8	8	8
Brazil	2	2	2	3	3	3	3	3	4	4	5	5	5	5	3	3
Canada	62	63	67	60	64	72	73	69	62	65	62	68	75	76	77	84
China	117	122	134	142	154	178	193	197	195	202	224	234	276	283	301	343
Colombia																
Czech Republic																
Denmark	10	11	12	12	11	12	12	10	10	14	15	12	13	14	13	15
Estonia																
Finland	3	3	3	2	2	2	2	2	2	2	4	3	3	2	2	2
France	16	18	19	20	19	21	19	20	21	26	29	30	32	31	31	29
Germany	21	27	28	28	26	25	25	24	24	26	25	26	26	28	25	26
Greece	2	2	3	3	3	3	3	2	2	2	2	2	2	1	1	1
Hungary	3	3	3	3	3	3	3	3	3	2	1	2	2	2	2	2
Iceland	1	1														
India	102	113	120	122	125	129	138	139	140	141	143	148	150	148	145	150
Indonesia	8	8	8	8	8	8	8	8	9	9	9	10	10	11	9	10
Ireland	5	7	7	8	8	9	8	10	11	13	15	14	13	16	16	16
Israel	16	19	21	21	22	25	26	28	29	35	35	35	33	34	36	33
Italy	2	5	7	9	8	7	7	7	7	6	7	7	7	7	7	7
Japan	56	56	60	62	67	70	70	72	72	74	74	76	78	78	77	86
Korea	84	84	89	89	93	94	98	97	104	117	126	149	159	188	198	214
Latvia	3	3	3	3	3	3	3	3	3	3	2	2	2	2	2	1
Lithuania	1	1	1	1	1	1	1	1								
Luxembourg												1	2	1	1	
Mexico				1	1	1	1	1	1	1	1	1	1	1	1	1
Netherlands	5	5	7	7	6	5	4	2	2	5	5	7	7	7	7	8
New Zealand	6	6	5	5	4	4	4	3	3	5	5	5	5	5	6	7
Norway	3	5	7	10	10	10	10	9	10	9	11	11	11	11	12	12
Portugal	1	1	1	1	1	1		1	1	1	1	1	1			
Romania	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Russia	3	4	6	7	7	8	7	8	8	8	8	8	6	2	2	2
Slovak Republic	1	1	1	1	1	1	1	1	1	2	2	1	1	1	1	1
Slovenia	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Spain	2	4	6	6	6	7	7	9	10	10	13	13	10	10	9	9
Sweden	19	22	25	26	25	27	29	29	26	28	31	34	38	57	53	60
Switzerland	13	13	14	14	15	15	16	16	16	15	15	18	19	21	21	20
United Kingdom	79	82	79	70	57	52	49	49	54	61	62	67	72	75	67	67
United States	389	406	422	385	353	340	337	339	366	409	434	448	471	505	536	600

Source: OECD analysis based on Refinitiv Datastream database

References

Bajgar, M. et al. (2020), "Coverage and representativeness of Orbis data", *OECD Science, Technology and Industry Working Papers*, No. 2020/06, OECD, Paris, https://www.oecd-ilibrary.org/docserver/c7bdaa03-en.pdf?expires=1610637428&id=id&accname=guest&checksum=53DC9F88B164890E05BFBF6CDA8D70AD (accessed on 14 January 2021). [3]

Copenhagen Economics (2018), *Study on the economic impact of supplementary protection certificates, pharmaceutical incentives and rewards in Europe*, European Commission, Brussels, https://doi.org/10.2873/886648. [6]

Damodaran, A. (n.d.), *Research and Development Expenses: Implications for Profitability Measurement and Valuation*, Stern School of Business, New York City, NY, http://www.stern.nyu.edu/~adamodar/pdfiles/papers/R&D.pdf. [5]

IFRS (2021), *IAS 38: Intangible Assets*, https://www.ifrs.org/content/dam/ifrs/publications/pdf-standards/english/2021/issued/part-a/ias-38-intangible-assets.pdf. [8]

Kalemli-Ozcan, S. et al. (2015), *How to Construct Nationally Representative Firm Level Data from the Orbis Global Database: New Facts and Aggregate Implications*, National Bureau of Economic Research, Cambridge, MA, https://doi.org/10.3386/w21558. [4]

KPMG (2021), *IFRS vs. US GAAP: R&D costs*, https://advisory.kpmg.us/articles/2017/ifrs-vs-us-gaap-rd-costs.html (accessed on 30 August 2021). [9]

OECD (2021), *Health at a Glance 2021: OECD Indicators*, OECD Publishing, Paris, https://doi.org/10.1787/ae3016b9-en. [10]

OECD (2021), *OECD R&D tax incentives database, 2021 edition*, OECD, https://www.oecd.org/sti/rd-tax-stats-database.pdf (accessed on 2 August 2023). [2]

OECD (2020), *Corporate Tax Statistics, Second Edition*, https://www.oecd.org/tax/tax-policy/corporate-tax-statistics-second-edition.pdf (accessed on 7 April 2021). [1]

OECD (2020), "The effects of R&D tax incentives and their role in the innovation policy mix: Findings from the OECD microBeRD project, 2016-19", OECD, Paris, https://www.oecd-ilibrary.org/science-and-technology/the-effects-of-r-d-tax-incentives-and-their-role-in-the-innovation-policy-mix_65234003-en (accessed on 7 April 2021). [11]

UN (2003), *NATIONAL ACCOUNTS: A PRACTICAL INTRODUCTION*, United Nations, Department of Economic and Social Affairs Statistics Division, https://unstats.un.org/unsd/publication/SeriesF/seriesF_85.pdf (accessed on 30 April 2021). [12]

Wong, C., K. Siah and A. Lo (2018), "Estimation of clinical trial success rates and related parameters", *Biostatistics* February, pp. 1-14, https://doi.org/10.1093/biostatistics/kxx069. [7]