

WORKING PAPER

# Modelling the possible returns to the NHS from private sector use of the 100K genomes database

**August 2019**

Finbarr Livesey  
Department of Politics and International Studies (POLIS)  
University of Cambridge

E: [tfl20@cam.ac.uk](mailto:tfl20@cam.ac.uk)

## Abstract

The development of publicly-owned genomics databases linked to health records challenges the standard model of drug development. The usual assumptions of transferring intellectual property (IP) or other knowledge from the public to the private sector do not hold. This article models the possible returns such a database as that developed through the 100K Genomes project might return to the UK healthcare system if those in control of the database are willing to engage in new contractual arrangements. This is possibly the only time the public sector has the advantage in negotiating these relationships and structures put in place now will remain for many years to come.

---

Publication from the Bennett Institute for Public Policy, Cambridge  
[www.bennettinstitute.cam.ac.uk](http://www.bennettinstitute.cam.ac.uk)

## Introduction

While there is much discussion on patient protections, confidentiality and other ethical issues surrounding the use of the large scale genomic database developed through the 100K Genomes project, there is currently little discussion on how the public sector might benefit financially from the use of the database by private sector actors.<sup>1,2</sup> This is a significant omission as this publicly created and publicly controlled resource is different in scale and scope to other resources which cross the boundary between the private and the public sector as part of 'traditional' drug development. The database that has been created, and is owned and controlled by a company wholly owned by the public sector, has significant potential value and will continue to grow in value as further sequences are added, heading towards a new target of five million genomes over the coming five years.<sup>3</sup>

A new drug target, a molecule, these are single items that can be developed further by private actors, once identified in publicly funded research, and then if successful offered for sale to the healthcare system here in the UK and globally. This is a gross over-simplification of the 'traditional' development pathway but given how long it has been established it is hard to step back and consider that there is the possibility of a different narrative, a different approach for the public sector in negotiating with pharmaceutical companies and others in developing new treatments.

Use of the 100K Genomes database is fundamentally different in that the database is an aggregation of thousands of individuals genetic sequences linked to their health records which companies essentially rent access to, along with use of an analytical environment in order to generate understanding and therefore value. This distinction means that there is a significant difference in the interaction and a much stronger position for the public sector to negotiate the terms under which returns may or may not be shared. At this point in time just using established or traditional norms on how this relationship should operate would be a significant error and would miss an opportunity to generate large scale funds for reuse within the healthcare system.

This article discusses why this is a particular moment for the NHS and Genomics England (GE) as they structure the commercial relationships with companies accessing the 100K Genomes database. The use of the database may include the development of new targets for therapies and diagnostics, as well as possibly repurposing existing drugs from one disease area into another, among other uses. The article suggests possible channels through which financial benefit could flow back to the healthcare system and provides a first pass model for one of the channels to have a ranging shot on the potential scale of such financial benefits.

## Why is this interaction between the public and private sector different?

The established narrative on the development of new drugs can be simplified to say that while the public sector does the early-stage high risk research, once there is a possibility that there is a molecule or a target that could be developed into a new treatment that should then be done in some private sector vehicle. For example, this may be through transferring the required intellectual property (IP) through a set of founders and a start-up or spin-out from a university. Or it may simply be the IP being licenced to a large existing pharmaceutical company who then drive the development in-house. Then after many years and significant investment the new drug or treatment is available for use having been through clinical trials and approval by NICE or the MHRA in the UK context. It then has a life based around patent protection in which it recoups those costs and eventually is profitable for the company.

This story has come under increasing challenge in recent years. Some of this is driven by the rates of price increases for prescription drugs<sup>4</sup>, with a number of high profile cases calling into question the claim that companies are pricing to recoup their costs and making a reasonable return, rather than attempting to price as high as the market will sustain (for example the case of the treatment provided by Gilead for hepatitis C in 2015).<sup>5</sup> There are also arguments on the actual costs of development. The cost level most quoted for the development of a new drug is in excess of \$2 billion based on ongoing work at the Tufts Center for the Study of Drug Development.<sup>6</sup> This is challenged by other studies which claim that the costs of development are much lower, of the order of \$650 million.<sup>7</sup> This paper is not focused on the costs of development and the arguments about what that allows or not in terms of drug pricing. However, it is important to contextualise the scale of investment that is in play, as this also then relates to the perceived value for the use of databases such as the 100K Genomes database and what it can do to reduce costs of development and to increase the potential return to private companies.

Beyond the challenge to the model of drug development itself and the role of the public and the private sector overall, this case of the 100K Genomes database is fundamentally different. The database has a number of possible uses which were initially characterised into three areas<sup>8</sup> –

- Diagnosis<sup>9</sup> - the use of the databases in diagnosis of the genetic basis of disease
- Pharmacogenomics - to work out which treatments are going to be more effective
- Developing new diagnostics and treatments - identifying genes and their role in disease can present new targets, as well as using the databases to see if existing drugs can be repurposed to address other diseases

The first is the focus for the new NHS Genomics Medicine Service<sup>10</sup> which intends to use the database to have faster diagnosis and better matching of treatments to patients so that survival rates specifically in cancer increase. This is not the focus of this paper as that is the NHS using the database that has been developed within the healthcare system, and while it will contact with ongoing research and development it is using the information in a more direct, immediate way.

For companies the second and third areas potentially allow them to reduce the time (and therefore costs) in development of a drug or treatment, provide them with new routes to drugs (repurposing), help identify targets with a higher likelihood of succeeding (i.e. making it through clinic and not failing late in the clinical trials pathway therefore reducing the number of failed developments). All of these are of significant value, as we have seen in terms of the costs of development whether it is of the order of \$600 million or anything up to and including \$2.6 billion. Companies will not use the database and its analysis environment unless it is useful and valuable to them. Understanding the scale of that potential value will be difficult and makes decision-making on both sides difficult at this point in time. However the expectation is that the database, the sequences, the linked health records and the analytical environment in which companies can make sense of the data, will provide significant value. Avoiding a development pathway that fails during phase III trials, shortening the time and investment required to move from pre-clinical to clinical trials, at these levels of investment it is clear that Genomics England have a resource of significant and increasing value that should be recognised and managed as such.

The public may already have a clear sense of the 100K Genomes database as something which should not be exploited without return to the public. A 2019 report sponsored by GE and ScienceWise completed a public dialogue project to understand how the public think about genomics and specifically the development of genomic medicine in the UK.<sup>11</sup> The report uses the metaphor of the social contract between those contributing to the database and those then using it. Whilst not a major part of the conversation there was a clear position on the use of the database by the private sector. “If the UK public are donating data, this brings in attitudes to the private sector and as such participants thought that the public has a right to expect that some of the financial rewards of new medical developments will return to ‘UK plc’. And therefore, participants thought that policy makers should implement a mechanism which delivers treatments for rare/ultra-diseases, a fair price for new treatments, and re-investment in clinical care and research, in recognition of the fact that genomics creates new financial opportunities for pharmaceutical companies.” Without having any discussion on costs of development or a deep understanding of the development process for new treatments, there is an immediate response from the public that this is a public resource that should not be exploited without the public at large benefitting.

The public position adopted on the Genomics England website appears to bias towards industry. The Discovery Forum is the first mode of interaction between companies and GE where it is claimed “Companies have come together within the Discovery Forum to work in a pre-competitive environment with access to a selection of whole genome sequences.”<sup>12</sup> Later in the FAQ on how industry works with GE is it stated with no justification that “Medicines and diagnostics are always developed outside the NHS and government by the private sector.”<sup>13</sup> This simplistic and explicit statement of the relationship appears to indicate that GE is not thinking of how to monetise or release the benefit from the database back to the public sector.

## What are possible channels for benefit to return to the healthcare system?

The usual approaches to intellectual property and licencing will not simply work for access to a database such as that created in the 100K Genomes project. As stated above there is no transfer of IP or of a specific licence to use a molecule, for example. Here companies are paying for access to a data resource in which they can carry out research work but take none of the original data out of the protected environment in which they do this work. How then can a link be made between companies who access the database and the returns they may see in the future from that access?

In a 2014 report for the Nuffield Bioethics Council there was early commentary on how this might be achieved. "A key assumption underpinning the whole project is that the results of the 100k Genome Project and other GeL activities will be commercialisable. GeL anticipate profit to arise from the dataset in several ways: i) Companies may simply pay to access GeL's database of sequences; ii) GeL as a company, may enter into a royalty sharing scheme or pursue a joint venture model, options designed to aid small and medium sized businesses; iii) companies may purchase exclusive time limited access to the dataset (Genomics England 2013d, p.8). These details are, at this time, still largely speculative."<sup>14</sup>

Since that point the original industrial trial with GE (the GENE consortium) has completed its work and companies are now making agreements with and working with GE to access and use the existing database. However, these are commercially sensitive so it is very hard to find details of what the emerging agreements contain.

Why is it important to have clarity on these possible mechanisms now? These are the early stages of the development of the use of large scale genomic databases, linked to healthcare records, for the development of new drugs or the repurposing of existing drugs into different disease areas. Even though it is early in this process, having a perspective of public benefit from the beginning is very important as the returns are going to be cumulative and the longer the public side of this interaction is not well formed, the smaller the return and impact that we would see. It is also possible that structures put in place now will exist far into the future and any disadvantage that the public sector places itself at will remain. It is not the job of the private sector to disadvantage themselves in what are commercial negotiations, GE is after all a company in its own right. However, there is a responsibility on GE to ensure that it is generating as much public benefit as possible given its founding and the fact that it is a wholly owned company of the Department of Health and Social Care (DHSC).

If we are thinking purely in financial terms, what might be the options for taking money back to the public from the private companies developing treatments using the database?

- The simplest approach is based on **access costs**, with companies either paying one time fees or annual membership for example. What do you charge for access to the data and more importantly the analytical framework so that the data can be translated into targets for development or other uses?

- Next we can think of GE contracting to take a percentage (a haircut) from the **profits or sales** for treatments hallmarked as developed using the database. Here companies would not have to pay up front to use the database (an incentive for smaller companies to become involved) but would only have to provide payments if a treatment was successful. Using sales may be a better approach than profits, as sales figures will be less liable to any accounting treatment or means to reduce their level. At its simplest level this would be set as a percentage of annual sales for the drug or treatment in question.
- Next GE could agree to take **equity** in companies that want to use the data. This would provide GE with a portfolio of investments that would need both to be well structured and managed over a long time period. Again this is more likely to be of interest or use in terms of start-ups or small companies rather than large established multinationals.
- In terms of the social contract spoken of in the GE report, another channel for direct return to the NHS would be to provide **free or discounted access** to treatments developed using the database. This would allow the companies to exploit the treatments developed worldwide without any further drag on returns, while the NHS would be able to reduce the large and increasing costs of prescriptions - according to the King's Fund the NHS prescription bill was £17.4 billion in 2016/2017.<sup>4</sup>

This is not intended to be an exhaustive or complete list. This is a set of possible options to which others can add into the future. However, it is a reasonable starting point when thinking of different approaches to generating return from the use of the database to the NHS in particular and the broader healthcare system depending on how such income might be used.

Two points should also be made at this point:

- The boundary for companies to exploit the developments based on the 100K database is not based on the UK. Companies can and will take the treatments or other benefits derived here and use them globally. Therefore negotiations should recognise the national versus global settings for exploitation.
- The database and its use will continue to increase in value over time, as the analytic environment is developed, further sequences are added (the intention now is to reach five million whole genome sequences in the next five years) and other types of sequences are added. Depending on how the future interaction is structured with the companies they will continue to realise value from their use of the database and the research they complete with it, that is there is ongoing value to the companies.

## **Modelling possible returns from sales**

As a first step in understanding the possible scale for financial return to GE and through GE the UK healthcare system the rest of this paper develops a model for the channel based on a percentage of sales for treatments developed using the database. This is based on a first assumption that it is possible to track which treatments have significantly used the database, either in developing new treatments, repurposing treatments from one area to another or otherwise significantly benefited for example through reducing by a long amount the time it takes to develop a treatment (i.e. a more indirect but still very significant benefit). It makes no other assumptions about whether GE uses access costs as well, it is an attempt to model just this channel of return.

The model itself is based on cancer drugs, given this is a very significant area for development and is one of the first elements that the GE database has significant data for and will continue to focus on. It is also restricted to cancer as there is good data on the approvals of such drugs aligned to their subsequent sales for companies. By modelling one channel we can provide some sense of a lower bound for the kinds of return that may be possible to achieve through managing the interface with companies using the database, if the model has a realistic underpinning.

In order to build the model of potential return a number of assumptions have to be made. Where possible we have used the best available number or data from previous research, from company data or from the regulator and the approvals of cancer drugs in the UK.

### **1. Number of cancer drugs coming to the market in the UK each year**

How many drugs does NICE approve each year? In cancer<sup>15</sup> it appears to be on average just under 14 per year since 2000, but currently a three year average of appraisals is at 40 as the number of appraisals has jumped significantly since 2016. The overall 'success rate' i.e. a positive recommendation is 73% according to the data from NICE.

### **2. How many of those will use the GE database?**

This is a key assumption which will depend heavily on how valuable the companies see access to the database and its analytic environment in order to either develop new targets, to aid in repurposing or to reduce the time and costs in development. As this is currently unknowable, and companies are under no obligation to clarify this point, this will be a variable that should be used as a range. It also assumes that we are able to monitor and enforce a link between the use of the database and the introduction at a later date of a successful cancer drug.

This variable will moderate how many drugs are then included in the portfolio of drugs from which benefit may flow back to GE and the NHS in general. We will use three values for what percentage of drugs will have used the GE database (in an abstract timeline in the future rather than in terms of the current levels of engagement). We will assume three possible levels of use at 2%, 5% and 10% of drugs coming to the UK market having been 'significantly helped' by the GE database and therefore have the basis for a commercial contract to return a share of sales to GE and the NHS.

### 3. What is the return in terms of sales for each drug?

How much revenue does each cancer drug generate per year? A recent study has looked at all of the cancer drugs approved by the Food and Drug Administration (FDA) from 1989 to 2017 in order to estimate cumulative and average sales per year for these drugs.<sup>16</sup> “The mean annual sales income since the year of approval ranged between \$3 million for Acabrutinib (approved in October 2017) and \$5.9 billion for Bevacizumab (approved in February 2004). **Thirty-three drugs (33.3%) were found to have mean annual sales income of more than \$1 billion.** [emphasis added]”<sup>16</sup> The drugs that earned the most had annual sales in excess of \$5 billion and these represent a small portion of what is referred to as the blockbuster end of the market.

With that information a coarse five point scale of annual yearly return, using the following values, can be developed (table 1). The lower end of the scale assumes that there is a return, even if minimal, to all of the drugs that have come to market, running from \$50 million per year through to the assumed blockbuster values for the top third of the distribution. These values can then be used on a probability basis for each drug that is approved.

Success level	Annual return (£million)
Very low	50
Below average	250
Average	600
Above average	1,000
Blockbuster	5,000

**Table 1 – input values for annual yearly return (£million)**



#### **4. What is the likelihood of a drug being at a given success level?**

Again using the Tay-Teo comments above, specifically that a third of the drugs have on average over £1 billion in sales, we can develop a simple five point probability distribution for what category each of the drugs that comes to market ends up in. The comment that a third earn more than \$1 billion anchors the top of the distribution, and the rest is a simple estimation placing the plurality of drugs in the average category, with one fifth of the drugs having below average earnings.

The probabilities used in the model are shown in table 2 and again can be adjusted depending on how further data becomes available to show the current distribution of success for cancer drugs in the UK and globally.

<b>Success level</b>	<b>Probability of being in the level stated (%)</b>
Very low	5
Below average	15
Average	47
Above average	30
Blockbuster	3

**Table 2 – Probability of a drug being at a level of success**

#### **5. What is a realistic level of haircut to take from the company's sales?**

This again is an unknowable quantity. It depends critically on the perceived value that companies see in using the database as well as how GE and other actors approach negotiating the terms of use for access and analytical support. One key element is if use of the database not only reduces the direct costs for the development of a successful treatment, but if it also allows companies to reduce the number of unsuccessful developments – for example avoiding one to two developments that reach a

late stage of development (for example phase II or phase III trials) costing them of the order of \$1 billion for no return.

The level of percentage taken from sales should be set so as to recognise this potential value to the companies but also at a level that encourages use rather than blocking companies from being able to remain revenue positive around the interaction. It should be noted however that these are payments on success rather than upfront and through development. The company using the GE database is only committed to paying GE for the use of the database if their development is successful. This is why an approach for GE that uses access costs to manage their basic infrastructure and analytical support costs combined with a channel such as this to manage the longer term return of benefit to the NHS is likely.

It is also likely that the pharmaceutical companies will downplay the importance of the use of the database, characterise it as early stage research or analysis, in order to position themselves to maintain the maximum return they can achieve. This narrative has to be resisted, as again this interaction is fundamentally different to the usual sharing or transfer of intellectual property across the public-private boundary in drug development.

As there is likely to be a range of possible positions that GE could take in negotiating the return to themselves and the NHS and a spectrum of values that will be acceptable to the companies involved, this again is a variable that we have used on a simple three point scale. The initial input values we have used in the model are 1%, 2% and 5% of yearly sales for successful drugs, to show the potential range that might be achieved over the long term.

## Output of the initial model

Bringing these assumptions together a model for how a portfolio of drugs and treatments may develop for companies that choose to use the GE database at some point in the future can be built. This is not a model of the current state of the interaction between companies and GE, rather it is intended to represent how a portfolio of development and benefit may exist at some point in the future.

The model runs in the following order –

1. Each year a number of cancer treatments are assumed to be submitted to NICE for possible approval for use, anchored in the first year at 30 approvals. This is then moderated by the stated success rate of 73% for NICE approvals.
2. Next this is reduced according to the number of treatments that have used the 100K Genomes database, according to how many have made significant use of the 100K database (point 2 in the section above).
3. Each drug is then assigned to a category of success using the probability distribution developed in point 4 above.
4. This then for that year gives a number of drugs and an average sales return based on the success category that each is assigned to when they enter the portfolio.

Once added, the drugs stay in the portfolio providing the same level of return to the portfolio each year. For example in year one, two drugs are added both in the average category. For that year the return would be the sum of those two drugs. In year two, three further drugs are added, one below average, one average and one above average. For this year the return will now be the sum of the five drugs, two from the first year and three from the second year.

One of the key points is that this is a cumulative benefit with an assumption that such aggregation will plateau around 20 years into the process. This is based on having a long period of exploitation for the drugs developed and as drugs are no longer used other drugs will come into the portfolio. It is assumed that average yearly returns will stabilise around a twenty year time horizon. Therefore the model was run assuming a twenty year time horizon, with no loss of drugs that were part of the portfolio. In reporting the results we provide the average yearly return that the model indicated at 10 years and at 20 years.

To have a meaningful result, the model was run 20 times with the assumed probabilities providing a different level of return in each run. The key assumptions in the model of how companies would interact with GE and use the database are contained in the percentage of successful drugs that come to market that have used the database, and the level of haircut that has been negotiated with the companies on a flat basis, that is that the same level of percentage of sales returned is assumed across all of the portfolio. The average, maximum and minimum of average yearly return at 10 and 20 years is given in tables 3 and 4 in £ million.

	2% of drugs; 1% of sales	2% of drugs; 2% of sales	2% of drugs; 5% of sales	5% of drugs; 1% of sales;	5% of drugs; 2% of sales	5% of drugs; 5% of sales	10% of drugs; 1% of sales	10% of drugs; 2% of sales	10% of drugs; 5% of sales
<b>Average</b>	86	183	414	188	316	890	361	708	1824
<b>Maximum</b>	47	113	258	118	224	568	240	567	1348
<b>Minimum</b>	151	333	775	308	458	1183	508	970	2263

**Table 3 –average yearly return at the ten year point based on 20 iterations (£ million)**

	2% of drugs; 1% of sales	2% of drugs; 2% of sales	2% of drugs; 5% of sales	5% of drugs; 1% of sales;	5% of drugs; 2% of sales	5% of drugs; 5% of sales	10% of drugs; 1% of sales	10% of drugs; 2% of sales	10% of drugs; 5% of sales
<b>Average</b>	259	521	1259	594	1109	2722	1132	2301	5943
<b>Maximum</b>	175	362	860	478	874	2195	883	2010	5285
<b>Minimum</b>	339	703	1913	727	1376	3570	1376	2616	7188

**Table 4 –average yearly return at the twenty year point based on 20 iterations (£ million)**

The results at twenty years are thought to be more representative of the potential returns (in current pounds) due to the cumulative nature of such a portfolio, with new drugs entering the portfolio over time and older drugs going out of use. The mid-range figure with a 2% haircut on sales and 5% of new cancer treatments coming to market in the UK each year having significant use of the database shows a return to GE and the healthcare system of over a billion pounds on average. For context the current total NHS budget is of the order of £120 billion, so this is approaching one percent of the spending on public healthcare in the UK.

## Discussion

The model developed here is a first attempt to estimate how financial benefits might flow from the use of the 100K Genomes database to the public sector. Due to the fundamentally different nature of the database, its public ownership and the nature of its use, established approaches to the private sector exploiting the outputs of publicly funded activity (specifically research and development) do not hold.

This model suggests that there is a very broad possible range of returns to the public sector based on companies using the 100K database. It is very important that the public sector is thinking long term in order to not lock in agreements that do not allow them to garner the benefits that they should for GE, the NHS and patients in the UK.

The scale of returns that this model suggests ranges from £250 million per year to just under £6 billion per year, with a reasonable mid-point of over £1 billion per year on average given the assumptions of the model. This paper is not seeking or claiming accuracy on what the actual returns will be as new drugs are developed. Rather the is attempting to highlight that the possible returns for the NHS are significant and how GE and the public sector are considering structuring agreements with private companies is critical. If there are average annual returns of £1 billion to the NHS this is of the order of 1% of the total NHS budget. With continuing pressures on public sector budgets and the likelihood that more investment will be needed due to rising levels of co-morbidity, changes in demographics and continued public pressure for more and better treatments, this could be a key source of funds for the NHS into the future.

Genomics England has to think differently to other agencies, both in terms of having a valuable asset that they will retain and will give them leverage in the market, and to situate themselves as developing its value over a period of twenty plus years rather than being in a one time negotiation with a company over a single target or molecule.

## References

- 1 Middleton, A. Society and personal genome data. *Human Molecular Genetics* **27**, R8 - R13 (2018).
- 2 Samuel, G. N. & Farsides, B. Public trust and 'ethics review' as a commodity the case of Genomics England Limited and the UK's 100,000 genomes project. *Medicine, Healthcare and Philosophy* **21**, 159 - 168 (2018).
- 3 Anonymous. Health secretary says 5 million genomes will be sequenced in next five years. *The Pharmaceutical Journal* doi:10.1211/PJ.2018.20205539 (2018).
- 4 Ewbank, L., Omojomolo, D., Sullivan, K. & McKenna, H. The rising costs of medicines to the NHS: What's the story?, (The King's Fund, London, 2018).
- 5 Nocera, J. *Wall Street wants the best patents, not the best drugs*, <https://www.bloomberg.com/opinion/articles/2018-11-27/gilead-s-cures-for-hepatitis-c-were-not-a-great-business-model> (2018).
- 6 DiMasi, J. A., Grabowski, H. G. & Hansen, R. W. Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of Health Economics*, 20 - 33 (2016).
- 7 Prasad, V. & Mailankody, S. Research and Development spending to bring a single cancer drug to market and revenues after approval. *JAMA Internal Medicine* **177**, 1569 - 1575 (2017).
- 8 Murray, L. & Border, P. (ed Parliamentary Office of Science and Technology) (POST, 2015).
- 9 Taylor, J. C. *et al.* Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nature Genetics* **47**, 717, doi:10.1038/ng.3304 <https://www.nature.com/articles/ng.3304#supplementary-information> (2015).
- 10 NHS England. *NHS Genomic Medicine Service*, <https://www.england.nhs.uk/genomics/nhs-genomic-med-service/> (Undated).
- 11 Ipsos Mori. A public dialogue on genomic medicine: time for a new social contract. (London, 2019).
- 12 Genomics England. *How we work with industry*, <https://www.genomicsengland.co.uk/working-with-industry/> (2018).
- 13 Genomics England. *FAQs about how we are working with industry*, <https://www.genomicsengland.co.uk/working-with-industry/working-with-industry-faqs/> (2018).
- 14 Martin, P. & Hollin, G. A new model of innovation in biomedicine?, (Nuffield Council on Bioethics, London, 2014).
- 15 NICE. *Technology appraisal data: cancer appraisal recommendations*, <https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance/data/cancer-appraisal-recommendations> (2019).
- 16 Tay-Teo, K., Ilbawi, A. & Hill, S. Comparison of sales income and research and development costs for FDA-approved cancer drugs sold by originator drug companies. *JAMA Network Open* **2**, doi:10.1001/jamanetworkopen.2018.6875 (2019).