



# Using Real-World Data to Predict Clinical and Economic Benefits of a Future Drug Based on its Target Product Profile

Christoph Gerlinger<sup>1,2</sup> · Thomas Evers<sup>3</sup> · Jeremy Rassen<sup>4</sup> · Richard Wyss<sup>5</sup>

Published online: 12 June 2020  
© The Author(s) 2020

## Abstract

**Introduction** For a new drug to be developed, the desired properties are described in a target product profile.

**Objective** We propose a framework for using real-world data to measure the disease-specific costs of the current standard of care and then to project the costs of the proposed new product for early data-driven portfolio decisions to select drug candidates for development.

**Methods** We sampled from a cohort of patients representing the current standard of care to generate a hypothetical cohort of patients that fits a given target product profile for a new (hypothetical) treatment. The healthcare costs were determined and compared between standard of care and the new treatment. The approach differed according to the number of outcomes defined in the target product profile, and the cases for one, two, and three outcome variables are described.

**Results** Based on assumed hypothetical treatment effect, absolute risk and cost reductions were estimated in a worked example. The median costs per day for one patient were estimated to be \$10.37 and \$8.39 in the original and hypothetical cohorts, respectively. This means that the assumed target product profile would result in cost savings of \$1.98 per day and patient—not accounting for any additional drug costs.

**Conclusions** We present a simple approach to assess the potential absolute clinical and economic benefit of a new drug based on real-world data and its target product profile. The approach allows for early data-driven portfolio decisions to select drug candidates based on their expected cost savings.

## Key Points

A framework to assess the potential benefit of a new drug based on real-world data and its target product profile.

The approach allows for early data driven portfolio decisions to select drug candidates based on their expected cost savings.

A worked example is included.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s40801-020-00203-w>) contains supplementary material, which is available to authorized users.

✉ Christoph Gerlinger  
christoph.gerlinger@bayer.com

<sup>1</sup> Statistics and Data Insights, Bayer AG, Müllerstraße 178, 13353 Berlin, Germany

<sup>2</sup> Gynecology, Obstetrics and Reproductive Medicine, University Medical School of Saarland, Homburg/Saar, Germany

## 1 Introduction

For a new drug to be developed, the desired properties are described in a target product profile [1]. This target product profile states the expected efficacy (“intended”) and safety (“unintended”) outcomes. For example, as compared with the current standard treatment, the target product profile could state that a reduction of 30% in efficacy outcome A (e.g., myocardial infarction) and of 20% in efficacy outcome B (e.g., ischemic stroke), at the potential risk of an increase of 20% in an adverse event C (e.g., gastrointestinal bleeding), is expected [2]. This is a typical benefit/risk trade-off and, for the purposes of this paper, we assume that this

<sup>3</sup> Global Market Access, Bayer AG, Wuppertal, Germany

<sup>4</sup> Aetion, Inc., New York, NY, USA

<sup>5</sup> Division of Pharmacoepidemiology and Pharmacoeconomics, Harvard Medical School, Boston, MA, USA

trade-off yields a net positive clinical effect for patients. This is, of course, an assumption that needs to be considered carefully and with input from all relevant stakeholders; should this assumption not hold, a wholesale reconsideration of the development program would be clearly warranted.

For payers—governments or commercial entities, those making formulary and other financial decisions—the question is, how does such a product profile translate into benefit and risk to patients, and how does it translate to costs for the healthcare system. On the cost question, payers ask whether the additional costs for the new drug and any incremental risk of adverse events will be offset by savings because of the drug's efficacy. This evaluation of costs is done in comparison with an existing standard of care, whose costs may be unknown but are generally measurable with existing data sources (e.g., healthcare claims). This cost evaluation is a critical step, as investment decisions in drug development are often (though not always) made in favor of drugs that offer clear advantages to existing treatments with respect to benefits, risks, and/or costs. Among payers' health technology assessment agencies and formulary decision makers, only drugs with clear benefit/risk and benefit/cost advantages will be positively viewed and ultimately accepted for reimbursement.

We propose a framework for using real-world data (RWD) to (1) measure the disease-specific costs of a disease's current standard of care and then (2) project the costs of the proposed new product, from which (3) a cost differential can be calculated. The approach employs retrospective RWD to obtain assumptions around the observed healthcare costs for a cohort of patients taking the existing standard of care. Recognizing that the new product may not be suitable for all patients using the current standard—for example, their disease may be less severe than that of the intended patients for the new medication, or they may be contraindicated for the new drug—the approach then takes a weighted sample from this baseline cohort to create a new cohort of patients who fit the target profile of the hypothetical drug under development and in whom hypothesized incremental effects on efficacy and safety can be estimated. This new sampled cohort and the baseline cohort representing the standard of care are then used to estimate differential disease-specific costs for the hypothetical treatment at both the population and the patient level.

The use of sampling and reweighting methods to generate a “pseudo-population” that reflects the characteristics of a target population has a rich history in the epidemiology and causal inference literature. In these settings, the objective is to generate a pseudo-population that reflects the covariate distributions of a target population to improve internal and external validity when estimating treatment effects (e.g., confounding control or generalizing study results) [3–10]. For these purposes, a large number of variables are often considered when reweighting the original cohort, which

is frequently achieved through various propensity score weighting/resampling approaches [7, 9–12].

With that said, the objectives for generating a pseudo-population for the purposes described here are separate from those for causal inference. Instead of generating a pseudo-population that mimics the covariate distributions of a target population, the sampling objectives here are simply to generate a pseudo-population where the expected number of outcome events reflects the expectations described in the target product profile for a new drug. For the purpose of simplicity, we chose to match the sampling within stratum of pseudo-population to the target treatment outcome.

Here, we describe a framework that samples from a cohort of patients representing the current standard of care to generate a hypothetical cohort of patients that fits a given target product profile. We then illustrate the sampling framework using an example based on RWD from patients with incident heart failure. With the proposed framework, we seek to address two audiences. For developers of medical products, we seek to create a framework for data-driven decision making on the allocation of investments in new products. For governmental and commercial payers, we seek to simplify cost-effectiveness analyses of a new drug compared with the standard of care and augment such analyses with assumptions drawn from actual patient experience. Compared with traditional cost-effectiveness analyses, the proposed framework is more flexible as it does not require comparison of pre- and post-treatment within the clinical trial. It further requires minimal assumptions on healthcare costs when assessing cost effectiveness as these are derived directly from RWD.

## 2 Methods

The sampling approach for generating the hypothetical cohort for which healthcare costs were determined and compared differed according to the number of outcomes defined in the target product profile. In the following, the cases for one, two, and three outcome variables are described.

### 2.1 Generating the Hypothetical Sample for One Outcome Variable

For one outcome variable, e.g., *A*, we observe the proportions in the first row of Table 1 from the real-world

**Table 1** Observed and hypothesized proportions real-world data for one outcome variable

Proportions	<i>A</i>	No <i>A</i>	Total
Observed	<i>a</i>	<i>b</i>	1
Hypothesized	0.7 * <i>a</i>	0.3 * <i>a</i> + <i>b</i>	1

**Table 2** Observed proportions real-world data two outcome variables

	A	No A	Total
B	$a$	$b$	$a+b$
No B	$c$	$d$	$c+d$
Total	$a+c$	$b+d$	1

**Table 3** Hypothesized proportions real-world data two outcome variables

	A	No A	Total
B	$e$	$f$	$0.8 * (a+b)$
No B	$g$	$h$	$0.2 * (a+b) + (c+d)$
Total	$0.7 * (a+c)$	$0.3 * (a+c) + (b+d)$	1

**Table 4** Example with a total count of 100

	A	No A	Total
B	$e * 100$	$f * 100$	6
No B	$g * 100$	$h * 100$	94
Total	8	92	100

population. Based on the target product profile, a reduction of 30% for outcome A is envisaged. Thus, a hypothesized sample as shown in the second row of Table 1 is needed. As all the individual cell proportions are known in this case, generating a sample with the hypothesized margins is straightforward and can be obtained by stratified sampling among the cases with outcome A and the cases without outcome A.

## 2.2 Generating the Hypothetical Sample for Two Outcome Variables

For two outcome variables, consider Table 2, which shows the distribution of the outcomes in the real-world population. Suppose that the hypothesized effects include a reduction of the rate of A by 30% and a reduction of the rate of B by 20%. To obtain the marginal distribution of the outcomes in a hypothetical population under these constraints, the hypothetical effect sizes are applied to the margins of Table 2 as illustrated in Table 3. Unlike the case of one outcome variable, the table cells  $e, f, g,$  and  $h$  can no longer be calculated from the margins because the matrix in Eq. (1) has only rank 3 and can therefore not be inverted. In other words, the solutions for the cell values for  $e, f, g,$  and  $h$  are not unique. For small total sample sizes, all possible solutions of Eq. (1) can be determined numerically. However, to get a unique solution, additional constraints need to be applied:

**Table 5** Solutions to example in Table 4 with a total count of 100

$e * 100$	$f * 100$	$g * 100$	$h * 100$
0	6	8	86
1	5	7	87
2	4	6	88
3	3	5	89
4	2	4	90
5	1	3	91
6	0	2	92

$$\begin{pmatrix} e \\ f \\ g \\ h \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0.8 * (a+b) \\ 0.7 * (a+c) \\ 0.3 * (a+c) + (b+d) \\ 0.2 * (a+b) + (c+d) \end{pmatrix} \quad (1)$$

For example, for a sample size of 100 and the margins given in Table 4, there are a total of seven solutions (i.e., possible cell count combinations). These solutions are given in Table 5. In this example, there is a wide range of the events: 0–6% for outcome A and 2–8% for outcome B. Since cost is strongly correlated with outcome events, it is apparent that any cost estimate would be heavily dependent on the particular solution of Eq. (1). Thus, we need to define an additional constraint on the cell frequencies  $e, f, g,$  and  $h$  in Table 3 to get a unique solution.

Huber [13] provides a way to identify unique cell frequencies by setting the odds ratio between the cell frequencies in the original population

$$OR := \frac{e/f}{g/h}, \quad (2)$$

as an additional constraint on the cell frequencies for the sampled population. Alternatively, one can of course fix any one of the four cell frequencies  $e, f, g,$  and  $h$  in Table 3. In the latter case, the calculation of the remaining three cell frequencies in Table 3 is straightforward, whereas the calculation given the odds ratio is more complicated and is laid out in “Appendix”.

Lastly, one could also assume that the occurrence of the two events are independent of each other. In this special case, the determination of the cell counts is straightforward as the odds ratio will be 1 by definition. However, in the case of independence of the two outcomes A and B, an easier way to obtain the cell frequencies is to distribute the margin counts in, for example, each row separately with the assumed column proportions. Needless to say, in the case of two efficacy outcomes, these are almost always correlated and not independent. Thus, the researcher must define a meaningful constraint, be it for the odds ratio or for one of the table cells.

**Table 6** Real-world data three outcome variables

		A	No A	Total
B	C	a	b	a + b
	No C	c	d	c + d
	Total B	a + c	b + d	a + b + c + d
No B	C	e	f	e + f
	No C	g	h	g + h
	Total no B	e + g	f + h	e + f + g + h
Total		a + c + e + g	b + d + f + h	a + b + c + d + e + f + g + h = 1

### 2.3 Generating the Hypothetical Sample for Three Outcome Variables

The next level is the case with three outcome variables. Consider Table 6 representing the real-world population. From Table 6, we get the following four summation equations:

$$a + c + e + g = o_A \quad \text{with } o_A \text{ the proportion of A} \quad (3)$$

$$a + b + e + f = o_B \quad \text{with } o_B \text{ the proportion of B} \quad (4)$$

$$a + b + c + d = o_C \quad \text{with } o_C \text{ the proportion of C} \quad (5)$$

$$a + b + c + d + e + f + g + h = 1. \quad (6)$$

Similar to the case of two outcome variables, there is no unique solution. Either several of the individual cell proportions need to be fixed or independence must be assumed in order to arrive at a unique solution. In our example, we assume independence of the safety outcome C from both efficacy outcomes A and B. Hence, the problem of defining the eight cell frequencies in Table 6 reduces to solving Table 2 for stratum C and for stratum not C independently.

## 2.4 Methods for the Worked Example

### 2.4.1 Data Source

The data source for this study was the Optum Clinformatics Data Mart (CDM), which is a US health insurance claims database that includes longitudinally linked patient records. All patients in the Optum CDM database who met the inclusion criteria were included in the study population. Data in the Optum CDM are collected in an observational manner such that the management of the patient is determined by the patient and the caregiver and represents care as it is provided in routine clinical practice.

### 2.4.2 Cohort

We generated a cohort of patients with incident heart failure using appropriate *International Classification of Diseases, Ninth Revision (ICD-9)* and *Tenth Revision (ICD-10)* codes. This subset of the Optum CDM database included all relevant parameters needed, including medical history, baseline characteristics, outcome A, outcome B, outcome C, and disease-specific costs. Subgroups were defined using the information from the medical history and baseline characteristics.

### 2.4.3 Hypothetical Sample

We developed software in the R programming environment for generating the hypothetical sample in the way described above. It also performed all analyses required to estimate absolute risk reductions (overall and in subgroups) and potential cost savings based on the risk reductions resulting from the hypothesized effects. If the assumptions in the target product profile are correct, the new derived database, i.e. original plus hypothetical sample, should mirror the claims database a couple of years after launch.

**Table 7** Patient counts by outcome

Outcome A	Outcome B	Outcome C	Frequency original	Proportion original	Frequency hypothetical	Proportion hypothetical
No	No	No	35,783	0.6171	39,086	0.6741
Yes	No	No	6307	0.1088	3679	0.0634
No	Yes	No	4761	0.0821	4177	0.0720
Yes	Yes	No	1799	0.0319	774	0.0133
No	No	Yes	4692	0.0809	5432	0.0937
Yes	No	Yes	2041	0.0352	2459	0.0424
No	Yes	Yes	1589	0.0274	1478	0.0255
Yes	Yes	Yes	1013	0.0175	900	0.0155

**Table 8** Risk and event reductions

	Risk original	Risk hypothetical	Risk difference	Events original	Events hypothetical	Event difference
Outcome A	0.192	0.135	− 0.05774	11,160	7812	− 3348
Outcome B	0.158	0.126	− 0.03161	9162	7329	− 1833
Outcome C	0.161	0.177	0.01611	9335	10,269	934

### 3 Results

The described methodology generated cell frequencies and proportions for the hypothetical cohort based on the respective numbers from the original cohort (see Table 7). Based on the assumed hypothetical treatment effect in this example—a reduced efficacy outcome A by 30%, reduced efficacy outcome B by 20%, and increased safety outcome C by 10%—the absolute risk and event reductions were estimated (see Table 8). The median costs per day for one patient were estimated to be \$10.37 and \$8.39 in the original and hypothetical cohort, respectively. This means that the assumed target product profile would result in cost savings of \$1.98 per day and patient—not accounting for any additional drug costs.

### 4 Discussion

For a new drug development, the desired properties of the new compound are described in a target product profile, which outlines the expected efficacy and safety outcomes. In this paper, we presented a framework to use RWD to determine how such a product profile would translate into cost savings for the healthcare system in order to make portfolio decisions. The approach estimates these cost savings by creating a hypothetical sample from RWD that is based on the target product profile.

Generating a unique hypothetical sample from RWD with no additional assumptions other than those provided in the target product profile is only possible in the case of one outcome variable. For two or more outcome variables, additional constraints or the assumption of independence between the outcome variables is required. While independence between two efficacy outcome variables would appear to be unrealistic, it can be envisaged that, for example, an efficacy and a safety variable are independent. That is, the occurrence of a side effect is unrelated to whether the drug is efficacious in a given patient. Our approach can easily be extended to analyze subgroups of the intended target population. This could, for example, help to predict potential randomized controlled trial subpopulations with a favorable relationship between anticipated absolute risk reduction and size of the respective subgroup. In combination with

available evidence of the value of the identified subgroups, the approach might help to justify the inclusion of the subgroups in the statistical analysis plan of a randomized controlled trial.

A limitation of our approach to estimating potential cost savings is that we require all relevant outcomes to be available in a real-world database. While this holds true for hard endpoints such as hospitalization or myocardial infarction, this is not always the case for softer endpoints such as pain. Further, when sampling for multiple outcomes, the sampling framework requires investigators to specify the relationship between the outcomes to get a unique solution. Decisions on what constraints to implement when specifying these relationships is subjective, and findings can be dependent on the constraints specified. Two solutions to this could be as follows: (1) applying an integer linear programming optimization to maximize the cost reduction and (2) providing a range of cost reductions by looping through all possible solutions and estimating the median along with the upper and lower bounds of the calculated cost reductions. However, these enhancements are beyond the scope of this manuscript, and we leave this to future research.

### 5 Conclusion

We have presented a simple approach to assessing the potential absolute clinical and economic benefit of a new drug based on RWD and its target product profile. The approach allows for early data-driven portfolio decisions to select drug candidates for development based on their expected cost savings. One application of the described approach is to assess the relative value of different subgroups, which may support evidence-driven decisions on portfolio candidates, research and development plans, and market access strategies. Potential future extensions of the methodology should be explored and may include RWD-based estimations of incremental cost-effectiveness ratios or indirect comparisons.

**Acknowledgements** The authors thank the anonymous reviewers for their very helpful comments that greatly improved this paper.



**Author Contributions** CG, TE, JR, and RW: designed the study; RW and JR: wrote the R code for the example; CG: drafted the manuscript; CG, TE, JR, and RW: contributed to the interpretation of the findings and the final manuscript. All authors approved the final version.

**Availability of Data and Material** The R code together with a description file is available as Supplementary Material. The material also contains code to generate synthetic input data as we are unable to publish the actual data used. The data that support the findings of this study are available from Aetion Inc., but restrictions apply to the availability of these data, which were used under license for the current study.

## Compliance with Ethical Standards

**Funding** This work was funded by Bayer AG. Bayer AG had no influence in the design of the study and collection, analysis, and interpretation of data or in writing the manuscript. The publication charges were paid by Bayer AG.

**Conflict of interest** CG and TE are full-time employees of Bayer AG. JR is a full-time employee of Aetion Inc. RW has no conflicts of interest that are directly relevant to the content of this article.

**Ethics Approval and Consent to Participate** No ethics approval nor patients' consent was required for our analysis of anonymized data.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

## Appendix

Given Table 4 with the marginal proportions the four equations derived from the known margins of and from the additional constraint on the odds ratio are as follows.

$$e + f = H \quad \text{with} \quad H := 0.8(a + b) \quad (7)$$

$$e + g = D \quad \text{with} \quad D := 0.7(a + c) \quad (8)$$

$$e + f + g + h = 1 \quad (9)$$

$$eh = Ofg \quad \text{with} \quad O := \text{OR} \quad (10)$$

Adding Eqs. (7) and (8) yields

$$f + g = H + D - 2e. \quad (11)$$

Multiplying Eqs. (7) and (8) yields

$$(e + f)(e + g) = HD = e^2 + ef + eg + fg \quad (12)$$

$$fg = HD - ef - eg - e^2 = HD - e(f + g) - e^2. \quad (13)$$

Using Eqs. (11) in (13) yields

$$fg = HD - e(H + D - 2e) - e^2 = HD - (H + D)e + e^2. \quad (14)$$

Multiplying Eq. (9) by  $e$  and substituting  $eh$  using Eq. (10) yields

$$e^2 + ef + eg + Ofg = e. \quad (15)$$

Substituting  $fg$  by Eq. (14) and  $(f + g)$  by Eq. (11) yields

$$e^2 + e(H + D - 2e) + O(HD - e(f + g) - e^2) = e \quad (16)$$

Substituting  $(f + g)$  by Eq. (11) yields

$$e^2 + e(H + D) - 2e^2 + OHD - Oe(H + D - 2e) - Oe^2 - e = 0 \quad (17)$$

$$e^2 + e(H + D) - 2e^2 + OHD - Oe(H + D) + 2Oe^2 - Oe^2 - e = 0 \quad (18)$$

$$(O - 1)e^2 + ((1 - O)(H + D) - 1)e + OHD = 0. \quad (19)$$

This quadratic equation in  $e$  has the possible solutions given by Eq. (20):

$$e_{1,2} = \frac{-((1 - O)(H + D) - 1) \pm \sqrt{((1 - O)(H + D) - 1)^2 - 4(O - 1)OHD}}{2(O - 1)}. \quad (20)$$

Given that  $e$  stands for a cell proportion, only solutions in the interval  $[0; 1]$  are valid solutions for our problem. The remaining unknowns  $f$ ,  $g$ , and  $h$  can then easily be derived from Eqs. (7), (8), and (9).

If we assume an odds ratio  $O$  of 2, say, in our example given in Table 6 with  $H=0.06$  and  $D=0.08$ , formula (20) yields [14]  $e_1 = -0.7331$  and  $e_2 = 0.0131$ , rounded to four decimal places. It follows that  $f=0.0469$ ,  $g=0.0669$ , and  $h=0.8731$ . In our example with a total sample size of 100, the best approximation is choosing cell counts  $e = 1$ ,  $f=5$ ,  $g=7$ , and  $h=87$  in Table 5, which corresponds to the second row of the possible solutions in Table 6.

An alternative to fixing the odds ratio as a fourth constraint on the table cells in 4 as outlined above, one could also fix one of the table cells. For example, adding the equation

$$e = E \quad \text{with} \quad E \text{ a fixed cell number}, \quad (21)$$

to the Eq. in (1) would yield a unique solution (if it exists, of course).

## References

1. Tyndal A, Du W, Breder C. Regulatory watch: the target product profile as a tool for regulatory communication: advantageous but underused. *Nat Rev Drug Discov.* 2017;16(3):156. <https://doi.org/10.1038/nrd.2016.264>.
2. US Food and Drug Administration. Guidance for industry and review staff target product profile—a strategic development process tool. 2007. <https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm080593.pdf>. Accessed 19 May 2019.
3. Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology.* 2000;11:561–70.
4. Stuart E, Cole S, Bradshaw C, Leaf P. The use of propensity scores to assess the generalizability of results from randomized trials. *J R Stat Soc Ser A.* 2001;174:369–86.
5. Stuart E, Bradshaw C, Leaf P. Assessing the generalizability of randomized trial results to target populations. *Prev Sci.* 2015;16:475–85.
6. Stuart E, Ackerman B, Westreich D. Generalizability of randomized trial results to target populations: design and analysis possibilities. *Res Soc Work Pract.* 2018;28:532–7.
7. Cole S, Stuart E. Generalizing evidence from randomized clinical trials to target populations: the ACTG 320 trial. *Am J Epidemiol.* 2010;172:107–15.
8. Cole S, Hernan M. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol.* 2008;168:656–64.
9. Westreich D, Edwards J, Lesko C, Stuart E, Cole S. Transportability of trial results using inverse odds of sampling weights. *Am J Epidemiol.* 2017;186:1010–4.
10. Dahabreh I, Robertson S, Tchetgen E, Stuart E, Hernan M. Generalizing causal inferences from individuals in randomized trials to all trial-eligible individuals. *Biometrics.* 2018. <https://doi.org/10.1111/biom.13009>.
11. Hansen B. Bias reduction in observational studies via prognosis scores. Technical report, Statistics Department, University of Michigan, Ann Arbor, Michigan (2006).
12. Wyss R, Hansen B, Ellis A, Gagne J, Desa IR, Glynn R, Sturmer T. The “dry-run” analysis: a method for evaluating risk scores for confounding control. *Am J Epidemiol.* 2017;185:842–52.
13. Huber W. <https://stats.stackexchange.com/users/919/whuber>, how to derive  $2 \times 2$  cell counts from contingency table margins and the odds ratio, URL (version: 2016-10-20): <https://stats.stackexchange.com/q/241394>. Accessed 08 Mar 2020.
14. <https://www.wolframalpha.com>. Accessed 08 Mar 2020.