

Bellabeat _Capstone_ Case Study

Bellabeat

Company summary

Urška Sršen and Sando Mur founded Bellabeat, a high-tech company that manufactures health-focused smart products. Sršen used her background as an artist to develop beautifully designed technology that informs and inspires women around the world. Collecting data on activity, sleep, stress, and reproductive health has allowed Bellabeat to empower women with knowledge about their own health and habits. Since it was founded in 2013, Bellabeat has grown rapidly and quickly positioned itself as a tech-driven wellness company for women.

Products

- ✚ **Bellabeat app:** The Bellabeat app provides users with health data related to their activity, sleep, stress, menstrual cycle, and mindfulness habits. This data can help users better understand their current habits and make healthy decisions. The Bellabeat app connects to their line of smart wellness products.
- ✚ **Leaf:** Bellabeat's classic wellness tracker can be worn as a bracelet, necklace, or clip. The Leaf tracker connects to the Bellabeat app to track activity, sleep, and stress.
- ✚ **Time:** This wellness watch combines the timeless look of a classic timepiece with smart technology to track user activity, sleep, and stress. The Time watch connects to the Bellabeat app to provide you with insights into your daily wellness.
- ✚ **Spring:** This is a water bottle that tracks daily water intake using smart technology to ensure that you are appropriately hydrated throughout the day. The spring bottle connects to the Bellabeat app to track your hydration levels.

Skateholders

- ✚ **Urška Sršen:** Bellabeat's cofounder and Chief Creative Officer.
- ✚ **Sando Mur:** Mathematician and Bellabeat's cofounder; key member of the Bellabeat executive team.
- ✚ **Bellabeat marketing analytics team:** A team of data analysts responsible for collecting, analyzing, and reporting data that helps guide Bellabeat's marketing strategy. You joined this team six months ago and have been busy learning about Bellabeat's mission and business goals — as well as how you, as a junior data analyst, can help Bellabeat achieve them.

Task as a Junior Data Analyst

Analyze smart device usage data in order to gain insight into how consumers use non-Bellabeat smart devices then provide recommendations to the Bellabeat marketing team for their upcoming campaigns.

1-Ask

1. What are some trends in smart device usage?
2. How could these trends apply to Bellabeat customers?
3. How could these trends help influence Bellabeat marketing strategy?

2-Prepare

2.1 Data Content

- Dataset is available at <https://www.kaggle.com/arashnic/fitbit> :FitBit Fitness Tracker Data (CC0: Public Domain.Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute_level output for physical activity, heart rate, and sleep monitoring.Dataset was generated by respondents to a distributed survey via Amazon Mechanical Turk between 03.12.2016-05.12.2016.
- The data is organized in both long and wide format.It has 18 csv files on different activities including daily steps, daily calories.

2.2 Limitation of the data

Thirty three people are not enough to represent to women population. As data was collected in a survey we have doubt about the integrity of the data

2.3- Does the data ROCCC?

- Remember, data is ROCCC when it is Reliable, Original, Comprehensive, Current, and Cited.
- The data is not reliable because it has only 33 people
- The data is not original since it comes from third party provider
- The data is comprehensive because the parameters match the products
- The data is not current it has more than 5 years.
- The data is not cited since it comes from third party.

3- Process

I will use R Studio to complete the project.

3.1 Data selected

- DailyActivity_merged.csv •SleepDay_merged.csv

3.2 Installing and loading packages and libraries

```
install.packages("tidyverse")
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'  
## (as 'lib' is unspecified)
```

```
install.packages("lubridate")
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'  
## (as 'lib' is unspecified)
```

```
install.packages("dplyr")
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'  
## (as 'lib' is unspecified)
```

```
install.packages("ggplot2")
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'  
## (as 'lib' is unspecified)
```

```
install.packages("readr")
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'  
## (as 'lib' is unspecified)
```

```
install.packages("tidyr")
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'  
## (as 'lib' is unspecified)
```

```
install.packages("skimr")
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'  
## (as 'lib' is unspecified)
```

```
install.packages("janitor")
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'  
## (as 'lib' is unspecified)
```

```
install.packages("here")
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'  
## (as 'lib' is unspecified)
```

```
install.packages("readr")
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'  
## (as 'lib' is unspecified)
```

```
library(tidyverse)
```

```
## — Attaching packages —
```

```
tidyverse 1.3.1 —
```

```
## ✓ ggplot2 3.3.5   ✓ purrr  0.3.4
## ✓ tibble 3.1.6   ✓ dplyr  1.0.8
## ✓ tidyr  1.2.0   ✓ stringr 1.4.0
## ✓ readr  2.1.2   ✓ forcats 0.5.1
```

```
## — Conflicts —————
tidyverse_conflicts() —
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()   masks stats::lag()
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union
```

```
library(dplyr)
library(ggplot2)
library(readr)
library(tidyr)
library(skimr)
library(janitor)
```

```
##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##   chisq.test, fisher.test
```

```
library(here)

## here() starts at /cloud/project

library(readr)
```

Importing cvs files

```
dailyActivity_merged <- read_csv("dailyActivity_merged.csv")
```

```
## Rows: 940 Columns: 15
## — Column specification
```

```
—
## Delimiter: ","
## chr (1): ActivityDate
```

```
## dbl (14): Id, TotalSteps, TotalDistance, TrackerDistance, LoggedActivitiesDi...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

sleepDay_merged <- read_csv("sleepDay_merged.csv")

## Rows: 413 Columns: 5
## — Column specification

```

```
—
## Delimiter: ","
## chr (1): SleepDay
## dbl (4): Id, TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

3.3- Data cleaning

Using head() function to display the first six records:

he

```
head(dailyActivity_merged)

## # A tibble: 6 × 15
##   Id ActivityDate TotalSteps TotalDistance TrackerDistance LoggedActivitie...
##   <dbl> <chr>         <dbl>         <dbl>         <dbl>         <dbl>
## 1 1.50e9 4/12/2016      13162          8.5           8.5           0
## 2 1.50e9 4/13/2016      10735          6.97          6.97          0
## 3 1.50e9 4/14/2016      10460          6.74          6.74          0
## 4 1.50e9 4/15/2016       9762          6.28          6.28          0
## 5 1.50e9 4/16/2016      12669          8.16          8.16          0
## 6 1.50e9 4/17/2016       9705          6.48          6.48          0
## # ... with 9 more variables: VeryActiveDistance <dbl>,
## # ModeratelyActiveDistance <dbl>, LightActiveDistance <dbl>,
## # SedentaryActiveDistance <dbl>, VeryActiveMinutes <dbl>,
## # FairlyActiveMinutes <dbl>, LightlyActiveMinutes <dbl>,
## # SedentaryMinutes <dbl>, Calories <dbl>
```

head(sleepDay_merged)

```
## # A tibble: 6 × 5
##   Id SleepDay      TotalSleepRecor... TotalMinutesAsl... TotalTimeInBed
##   <dbl> <chr>         <dbl>         <dbl>         <dbl>
## 1 1503960366 4/12/2016 12:00:0...      1           327          346
## 2 1503960366 4/13/2016 12:00:0...      2           384          407
```

## 3	1503960366	4/15/2016 12:00:0...	1	412	442
## 4	1503960366	4/16/2016 12:00:0...	2	340	367
## 5	1503960366	4/17/2016 12:00:0...	1	700	712
## 6	1503960366	4/19/2016 12:00:0...	1	304	320

Display all the columns using colnames()function:

```
colnames(dailyActivity_merged)
```

```
## [1] "Id"                "ActivityDate"
## [3] "TotalSteps"        "TotalDistance"
## [5] "TrackerDistance"   "LoggedActivitiesDistance"
## [7] "VeryActiveDistance" "ModeratelyActiveDistance"
## [9] "LightActiveDistance" "SedentaryActiveDistance"
## [11] "VeryActiveMinutes" "FairlyActiveMinutes"
## [13] "LightlyActiveMinutes" "SedentaryMinutes"
## [15] "Calories"
```

```
colnames(sleepDay_merged)
```

```
## [1] "Id"                "SleepDay"        "TotalSleepRecords"
## [4] "TotalMinutesAsleep" "TotalTimeInBed"
```

Identify the structure of the data using str() function:

```
str(sleepDay_merged)
```

```
## spec_tbl_df [413 × 5] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ Id          : num [1:413] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
## $ SleepDay    : chr [1:413] "4/12/2016 12:00:00 AM" "4/13/2016 12:00:00 AM"
## "4/15/2016 12:00:00 AM" "4/16/2016 12:00:00 AM" ...
## $ TotalSleepRecords : num [1:413] 1 2 1 2 1 1 1 1 1 1 ...
## $ TotalMinutesAsleep: num [1:413] 327 384 412 340 700 304 360 325 361 430 ...
## $ TotalTimeInBed    : num [1:413] 346 407 442 367 712 320 377 364 384 449 ...
## - attr(*, "spec")=
## .. cols(
## .. Id = col_double(),
## .. SleepDay = col_character(),
## .. TotalSleepRecords = col_double(),
## .. TotalMinutesAsleep = col_double(),
## .. TotalTimeInBed = col_double()
## .. )
## - attr(*, "problems")=<externalptr>
```

```
str(dailyActivity_merged)
```

```
## spec_tbl_df [940 × 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ Id          : num [1:940] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
## $ ActivityDate : chr [1:940] "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...
```

```
## $ TotalSteps      : num [1:940] 13162 10735 10460 9762 12669 ...
## $ TotalDistance   : num [1:940] 8.5 6.97 6.74 6.28 8.16 ...
## $ TrackerDistance : num [1:940] 8.5 6.97 6.74 6.28 8.16 ...
## $ LoggedActivitiesDistance: num [1:940] 0 0 0 0 0 0 0 0 0 ...
## $ VeryActiveDistance : num [1:940] 1.88 1.57 2.44 2.14 2.71 ...
## $ ModeratelyActiveDistance: num [1:940] 0.55 0.69 0.4 1.26 0.41 ...
## $ LightActiveDistance : num [1:940] 6.06 4.71 3.91 2.83 5.04 ...
## $ SedentaryActiveDistance : num [1:940] 0 0 0 0 0 0 0 0 0 ...
## $ VeryActiveMinutes : num [1:940] 25 21 30 29 36 38 42 50 28 19 ...
## $ FairlyActiveMinutes : num [1:940] 13 19 11 34 10 20 16 31 12 8 ...
## $ LightlyActiveMinutes : num [1:940] 328 217 181 209 221 164 233 264 205 211 ...
## $ SedentaryMinutes : num [1:940] 728 776 1218 726 773 ...
## $ Calories : num [1:940] 1985 1797 1776 1745 1863 ...
## - attr(*, "spec")=
## .. cols(
## .. Id = col_double(),
## .. ActivityDate = col_character(),
## .. TotalSteps = col_double(),
## .. TotalDistance = col_double(),
## .. TrackerDistance = col_double(),
## .. LoggedActivitiesDistance = col_double(),
## .. VeryActiveDistance = col_double(),
## .. ModeratelyActiveDistance = col_double(),
## .. LightActiveDistance = col_double(),
## .. SedentaryActiveDistance = col_double(),
## .. VeryActiveMinutes = col_double(),
## .. FairlyActiveMinutes = col_double(),
## .. LightlyActiveMinutes = col_double(),
## .. SedentaryMinutes = col_double(),
## .. Calories = col_double()
## .. )
## - attr(*, "problems")=<externalptr>
```

Identify total of duplications using sum() function:

```
sum(duplicated(sleepDay_merged))

## [1] 3

sum(duplicated(dailyActivity_merged))

## [1] 0
```

Remove duplications using distinct()function:

```
sleepDay_merged <- unique(sleepDay_merged)
sleepDay_merged%>%
```

```
duplicated()%>%
sum()

## [1] 0
```

check for distinct value using sum(duplicated)) function:

```
sum(duplicated(dailyActivity_merged))

## [1] 0

sum(duplicated(sleepDay_merged))

## [1] 0
```

4.-Analyze

```
summary(dailyActivity_merged)

##      Id      ActivityDate      TotalSteps  TotalDistance
## Min.   :1.504e+09 Length:940      Min.    : 0 Min.    : 0.000
## 1st Qu.:2.320e+09 Class:character 1st Qu.: 3790 1st Qu.: 2.620
## Median :4.445e+09 Mode :character Median : 7406 Median : 5.245
## Mean   :4.855e+09          Mean   : 7638 Mean   : 5.490
## 3rd Qu.:6.962e+09          3rd Qu.:10727 3rd Qu.: 7.713
## Max.   :8.878e+09          Max.   :36019 Max.   :28.030
## TrackerDistance LoggedActivitiesDistance VeryActiveDistance
## Min.    : 0.000 Min.    :0.0000      Min.    : 0.000
## 1st Qu.: 2.620 1st Qu.:0.0000      1st Qu.: 0.000
## Median : 5.245 Median :0.0000      Median : 0.210
## Mean   : 5.475 Mean   :0.1082      Mean   : 1.503
## 3rd Qu.: 7.710 3rd Qu.:0.0000      3rd Qu.: 2.053
## Max.   :28.030 Max.   :4.9421      Max.   :21.920
## ModeratelyActiveDistance LightActiveDistance SedentaryActiveDistance
## Min.    :0.0000      Min.    : 0.000 Min.    :0.000000
## 1st Qu.:0.0000      1st Qu.: 1.945 1st Qu.:0.000000
## Median :0.2400      Median : 3.365 Median :0.000000
## Mean   :0.5675      Mean   : 3.341 Mean   :0.001606
## 3rd Qu.:0.8000      3rd Qu.: 4.782 3rd Qu.:0.000000
## Max.   :6.4800      Max.   :10.710 Max.   :0.110000
## VeryActiveMinutes FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes
## Min.    : 0.00 Min.    : 0.00 Min.    : 0.0 Min.    : 0.0
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.:127.0 1st Qu.: 729.8
## Median : 4.00 Median : 6.00 Median :199.0 Median :1057.5
## Mean   :21.16 Mean   :13.56 Mean   :192.8 Mean   : 991.2
## 3rd Qu.:32.00 3rd Qu.:19.00 3rd Qu.:264.0 3rd Qu.:1229.5
## Max.   :210.00 Max.   :143.00 Max.   :518.0 Max.   :1440.0
##      Calories
## Min.    : 0
```



```
## 1st Qu.:1828
## Median :2134
## Mean :2304
## 3rd Qu.:2793
## Max. :4900
```

```
summary(sleepDay_merged)
```

```
##      Id      SleepDay      TotalSleepRecords TotalMinutesAsleep
## Min. :1.504e+09 Length:410      Min. :1.00      Min. : 58.0
## 1st Qu.:3.977e+09 Class :character 1st Qu.:1.00      1st Qu.:361.0
## Median :4.703e+09 Mode :character Median :1.00      Median :432.5
## Mean :4.995e+09          Mean :1.12      Mean :419.2
## 3rd Qu.:6.962e+09          3rd Qu.:1.00      3rd Qu.:490.0
## Max. :8.792e+09          Max. :3.00      Max. :796.0
## TotalTimeInBed
## Min. : 61.0
## 1st Qu.:403.8
## Median :463.0
## Mean :458.5
## 3rd Qu.:526.0
## Max. :961.0
```

visualize

```
library(ggplot2)
```

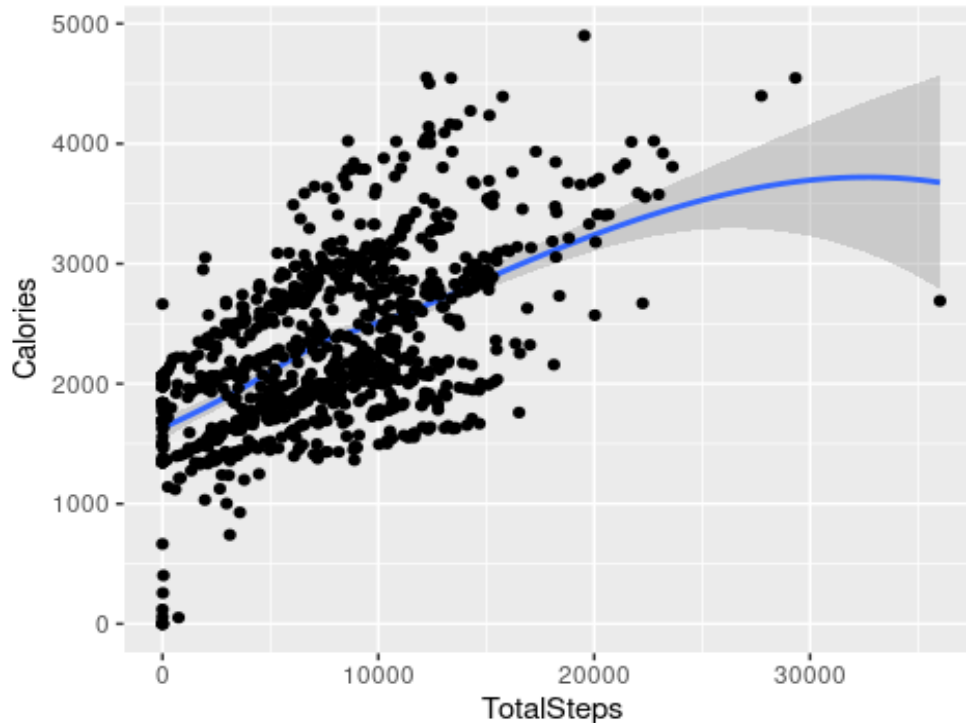
```
dailyActivity_merged
```

```
## # A tibble: 940 × 15
```

```
##      Id ActivityDate TotalSteps TotalDistance TrackerDistance LoggedActivitie...
##      <dbl> <chr>          <dbl>          <dbl>          <dbl>          <dbl>
## 1 1.50e9 4/12/2016      13162          8.5           8.5           0
## 2 1.50e9 4/13/2016      10735          6.97          6.97          0
## 3 1.50e9 4/14/2016      10460          6.74          6.74          0
## 4 1.50e9 4/15/2016      9762           6.28          6.28          0
## 5 1.50e9 4/16/2016      12669          8.16          8.16          0
## 6 1.50e9 4/17/2016      9705           6.48          6.48          0
## 7 1.50e9 4/18/2016      13019          8.59          8.59          0
## 8 1.50e9 4/19/2016      15506          9.88          9.88          0
## 9 1.50e9 4/20/2016      10544          6.68          6.68          0
## 10 1.50e9 4/21/2016      9819           6.34          6.34          0
## # ... with 930 more rows, and 9 more variables: VeryActiveDistance <dbl>,
## # ModeratelyActiveDistance <dbl>, LightActiveDistance <dbl>,
## # SedentaryActiveDistance <dbl>, VeryActiveMinutes <dbl>,
## # FairlyActiveMinutes <dbl>, LightlyActiveMinutes <dbl>,
## # SedentaryMinutes <dbl>, Calories <dbl>
```

5-Share

```
ggplot(data = dailyActivity_merged) + geom_smooth(mapping = aes(x=TotalSteps,y=Calories))  
+  
geom_point(mapping = aes(x=TotalSteps,y=Calories))  
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

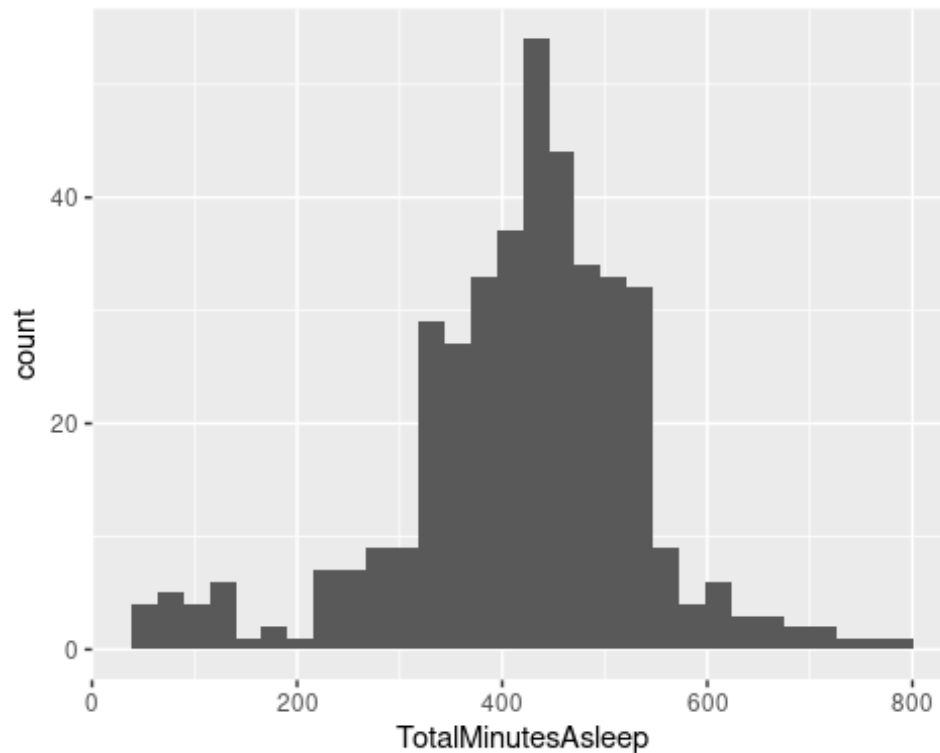


There is a positive correlation between the total steps and calories burned. When you increase the number of steps, you burn higher number of calories.

Based on [healthline.com](https://www.healthline.com/health/calories) Every day, you burn calories when you move around, exercise, and go about your daily tasks.

Most female adults need 1,600–2,200 calories per day, while adult males need 2,200–3,000 calories per day. However, the amount of calories you need each day is unique to your body and activity levels

```
ggplot(data = sleepDay_merged) + geom_histogram(mapping = aes(x=TotalMinutesAsleep))  
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Based on National sleep foundation in US adults from 25 to 65 years old should get 420 to 540 minutes of sleep daily. It is equal 7 to 9 hours of sleep everyday. We can see in the histogram above a high number of people sleep from 400 minutes.

6-Act

Recomendations

- ✚ Bellabeat can put notification on the app to encourage users to go to bed earlier.
- ✚ Create a platform as meetup where users ask questions and share stories using the products and goals in fitness.
- ✚ Bellabeat can create programs on the app how user can burn calories.

References

- ✚ <https://www.kaggle.com/arashnic/fitbit>
- ✚ <https://www.healthline.com/health/fitness-exercise/how-many-calories-do-i-burn-a-day#how-to-calculate>
- ✚ <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6267703/>