

## TABLE OF CONTENTS

---

1. Summary
2. Ask
  - a. Business Task
  - b. Stakeholders
3. Prepare
  - a. Data Source and Structure
  - b. Data Credibility and Integrity
  - c. Ethics and Compliance
  - d. Relevance and Limitations
4. Process
5. Analyze
  - a. Key Analysis Questions
  - b. Factor Tables
6. Share
  - a. Visualizations
    - i. Steps vs Calories Burned
    - ii. Sleep Duration vs Time in Bed
    - iii. User Distribution by Sleep and Activity Level
    - iv. Daytime Sedentary vs Wake Time in Bed
    - v. Activity Level by Sedentary Risk
    - vi. Step Count and Sleep Efficiency Histograms
7. Act
  - a. Business Recommendations
  - b. Future Considerations
8. Works Cited

## 1. Summary

Bellabeat is a high-tech company that designs health-focused smart products for women. Their devices track key wellness metrics, including physical activity, stress, sleep, menstrual cycle, and mindfulness, through the Bellabeat app. The company's mission is to empower women with accessible, beautifully designed technology that supports informed decisions about their health and daily habits. As a growing player in the smart device market, Bellabeat aims to strengthen its position and support further business growth by leveraging consumer insights.

## 2. Ask

- a. **Business Task:** Analyze smart device usage data from non-Bellabeat consumers to identify user behavior patterns and apply to Bellabeat's digital marketing strategies
- b. **Stakeholders:**
  - i. Urška Sršen - Bellabeat's cofounder and Chief Creative Officer
  - ii. Sando Mur - Mathematician and Bellabeat's cofounder and key member of the Bellabeat executive team
  - iii. Bellabeat Marketing Analytics team

## 3. Prepare

- a. **Data Source and Structure**
  - i. The dataset used in this analysis is the **Fitbit Fitness Tracker Data**, made publicly available through Mobius and hosted on Kaggle.
  - ii. The dataset includes 18 csv files containing various metrics from 33 fitbit users including heart rate, sleep monitors, hourly step count, calories, etc from 03.12.2016–05.12.2016.
  - iii. Stored in long format where each row represents a single observation at the minute level. A single ID has multiple rows with tracker data across the time span.
  - iv. For this analysis, I selected the following key files for relevancy:
    - 1. **dailyActivity\_merged.csv**
    - 2. **sleepDay\_merged.csv**

- b. **Data Credibility and Integrity**

Secondary data collected by a third-party individual that has limited documentation regarding the original data collection process. We do not know how the data donor selected participants or ensured data quality which may affect its overall credibility.

### c. Ethics and Compliance

The dataset was published under the [CC0: Public Domain license](#) on Kaggle, which permits unrestricted use, including for educational and commercial purposes. According to the source, all participants consented to share their anonymized data. There are no known privacy, licensing, or accessibility concerns associated with its use in this project.

### d. Relevance and Limitations

While the dataset offers detailed and comprehensive minute-level insights into user activity and sleep behavior, it presents several limitations in the context of Bellabeat's business needs:

- Demographic bias: Lacks demographic variables (gender, age, geographic location). Thus, it is difficult to assess whether the participants align with Bellabeat's target market of health-conscious women.
- Outdated data: Dataset was collected in 2016, so it's not current.
- Sample size: Very small with only 30 users included, limiting the generalizability of the findings
- Mismatch in features: Certain features relevant to our products and our customer segment are not included in this dataset including stress levels and menstrual cycle. This could result in lack of depth in analysis on relevant wellness dimensions.

Despite these limitations, the dataset still offers valuable and high-level behavioral insights on health-focused smart devices that can help guide Bellabeat's marketing strategy and inform assumptions about user engagement.

## 4. Process

**Tools Used:** RStudio, R

**Packages Used:** tidyverse, dplyr, ggplot2, janitor, lubridate, tidyr

### Preparation Steps:

1. Load necessary libraries and relevant datasets
2. Initial data inspection
  - a. `str()` to view basic dataset information (data types, row numbers, etc)
  - b. `head()` to preview the first few observations structure and data types
3. Clean each dataset
  - a. `clean_name()` for consistent formatting for column names
  - b. `is.na()` to view missing values
  - c. `distinct()` to keep only unique rows
  - d. `mutate()` to correct date column to correct data type

4. Merge datasets on ID and date column
5. Quality Checks

## 5. Analyze

### A. Key Analysis Questions

- a. Activity-focused - Leaf tracks activity and calories burned:
  - i. What is the average number of daily steps and active minutes among users?
  - ii. How much time do users spend in sedentary vs very active minutes?
  - iii. What is the relationship between step count and calorie expenditure?
- b. Sleep-focused - Leaf tracks sleep:
  - i. What is the average duration of sleep among users?
  - ii. How efficient is sleep (comparing total minutes asleep and total time in bed)?
  - iii. Are users getting the recommended hours of sleep (7-8 hours)? It is actually higher for women.
- c. Combined Activity and Sleep
  - i. Does more sleep per night correlate to higher or lower activity levels?
  - ii. How does physical activity vary by sleep duration? What is the average step count for users <6 hours, 6-8 hours, and more than 8 hours of sleep per night?
  - iii. Is there a pattern between bedtime efficiency and sedentary minutes?

### B. Analysis/Transformation Process

1. Added sleep efficiency column to sleep dataset
2. Viewed summary statistics for sleep and combined sleep/activity
3. Calculated daily average metrics per user
4. Created categories among users based on activity, sleep, and sedentary levels
5. Converted categories to factor levels

### C. Factor Tables

For **activity levels**, users were divided into categories based on their average daily steps.

Activity Level	Average Daily Steps
Sedentary	Less than 5,000
Low Active	5,000 to 7,500
Somewhat Active	7,500 to 9,999
Active	More than 9,999

\* Categories were referenced according to a [publication from the National Library of Medicine](#).

For **sleep levels** users were divided into groups based on whether they meet the recommended hours of sleep each time.

Sleep Recommendation Level	Average Hours of Sleep
Below	Less than 7 hours
Meets	7 to 9 hours
Above	More than 9 hours

\* Categories reference adult recommended hours of sleep each night in [publication from the National Library of Medicine](#).

Lastly, users were categorized into **sedentary levels** based on the amount of time spent being inactive, reflecting varying degrees of associated health risk.

Sedentary Risk Levels	Average Sedentary Time
Low	Less than 4 hours
Medium	4 to 8 hours
High	8 to 11 hours
Very High	More than 11 hours

\* These sedentary categories were based on guidelines from the 'Just Stand' campaign, as highlighted in a [Medical News Today article](#).

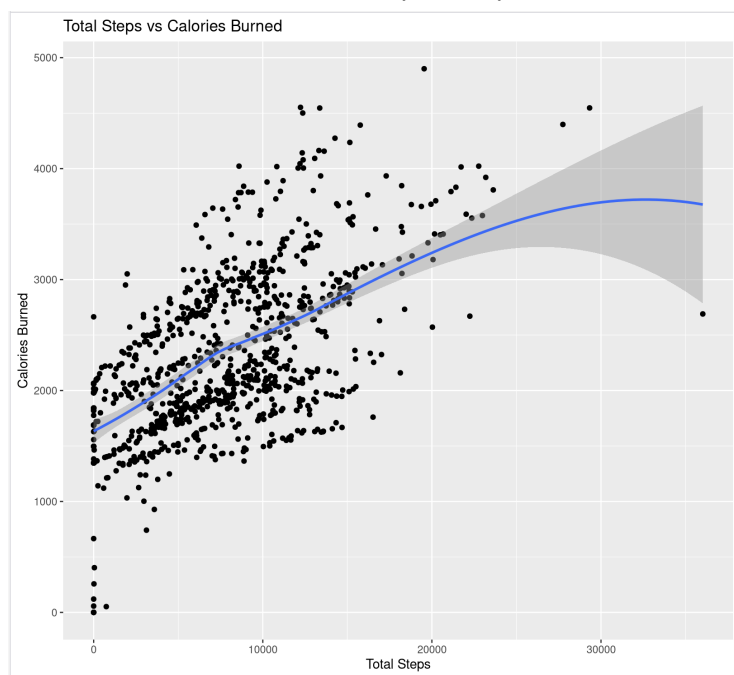
## 1. Share

### Exploratory Visualizations

- **Steps vs. Calories Burned:** Shows the relationship between physical activity and energy expenditure.
- **Sleep Duration vs. Time in Bed:** Highlights sleep efficiency, with a red-dashed reference line indicating perfect efficiency (every minute in bed was spent asleep).
- **Sleep Level by Activity Level:** Proportions of sleep categories across activity groups.
- **Daytime Sedentary vs. Wake Time in Bed:** Explores whether more sedentary time is linked to less restful sleep.
- **Activity Level by Sedentary Risk:** Compares activity levels across different sedentary categories.
- **Step Count and Sleep Efficiency Histogram:** Illustrates variability in physical activity and sleep quality over time across all recorded observations.

**Steps vs. Calories Burned**

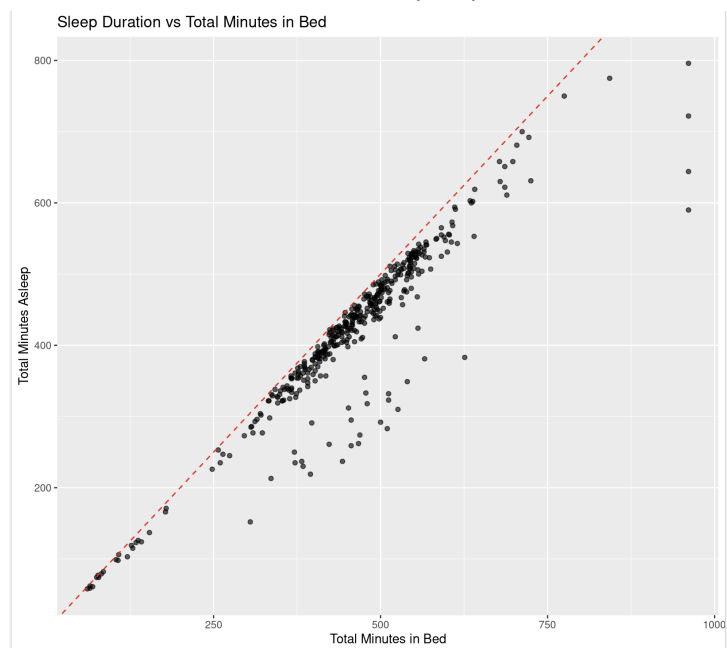
Dataset: Daily Activity

**Insights**

- ❖ There is a positive correlation between Total Steps and Calories burned: the more steps taken, the more calories are burned.

**Sleep Duration vs. Time in Bed**

Dataset: Sleep Day

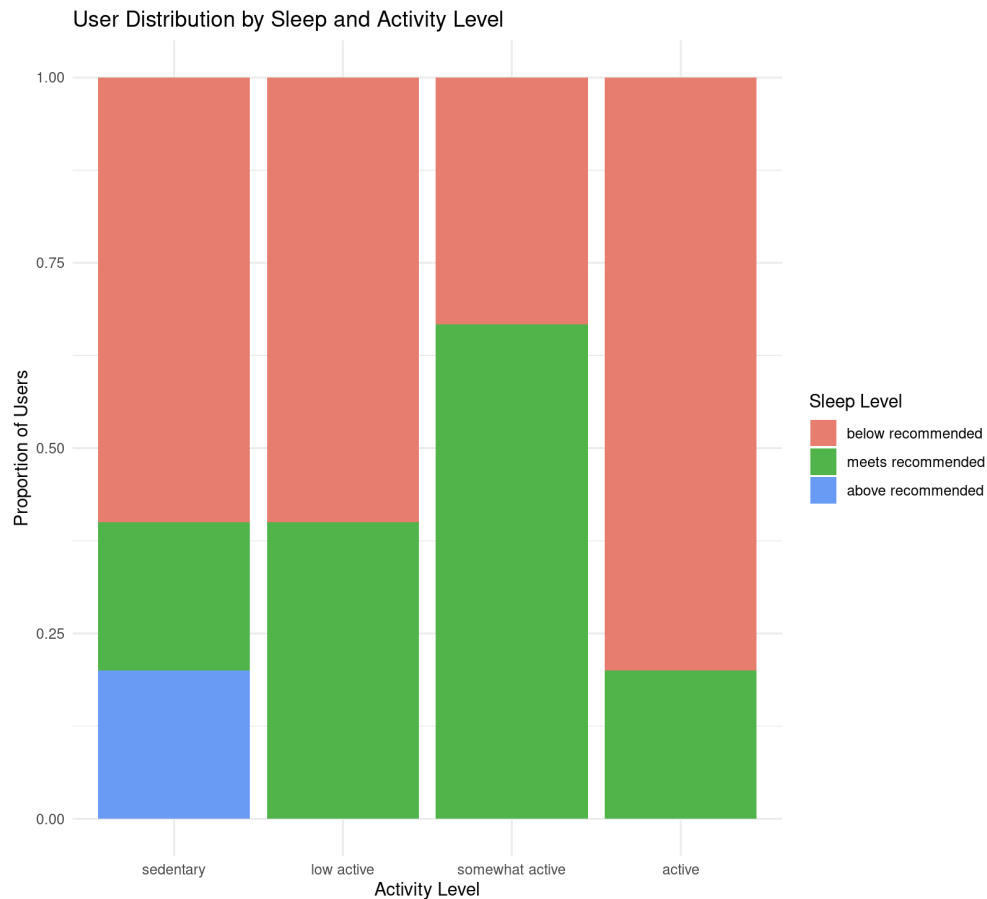


### Insights

- ❖ The dataset has a slope of 0.866 and an intercept of about 22. For every 1 minute spent in bed, users on average sleep 0.866 minutes.
- ❖ About 13.4% of time in bed is spent not sleeping.
- ❖ Compared to perfect efficiency, users overall display lower sleep efficiency.

### Sleep Level by Activity Level

Dataset: Combined Categories (Daily Activity + Sleep)



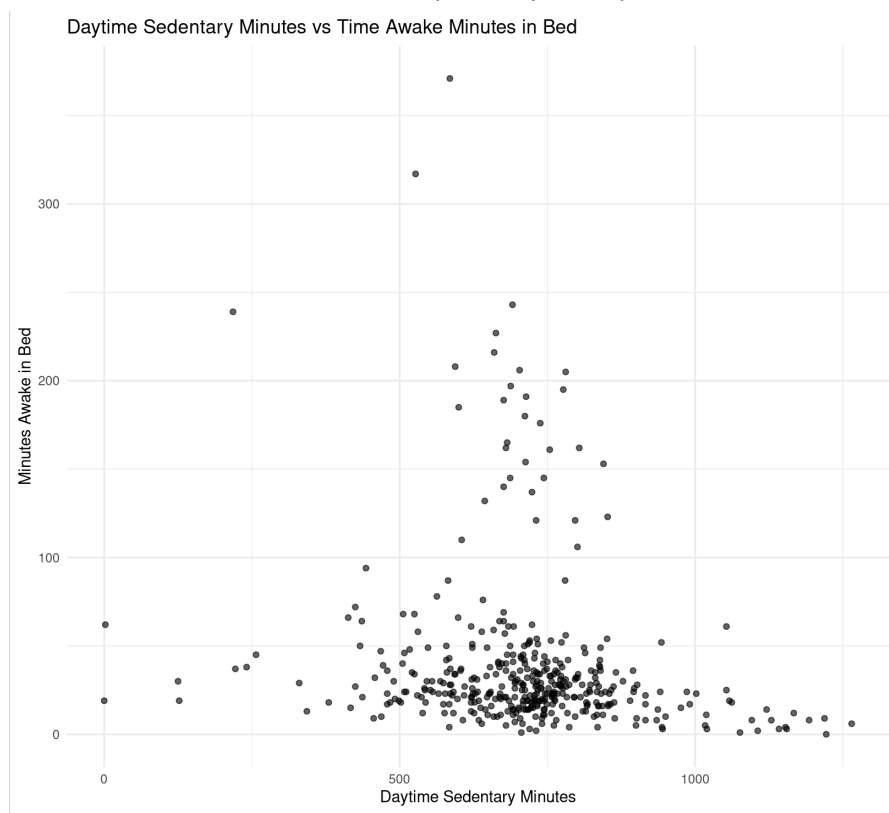
### Insights

- ❖ Over half of users fell into the below recommended sleep category.
- ❖ The somewhat active group had the highest proportion of users meeting the recommended sleep duration.
- ❖ In the active group, the mean sleep duration is 361 minutes or about 6 hours. 80% of users averaged less than the recommended sleep.

Note: The dataset is limited to 24 users with sleep data, so observations should be interpreted with caution due to small sample size and may not reflect trend to a larger population.

**Daytime Sedentary vs. Wake Time in Bed:**

Dataset: Daily Activity + Sleep

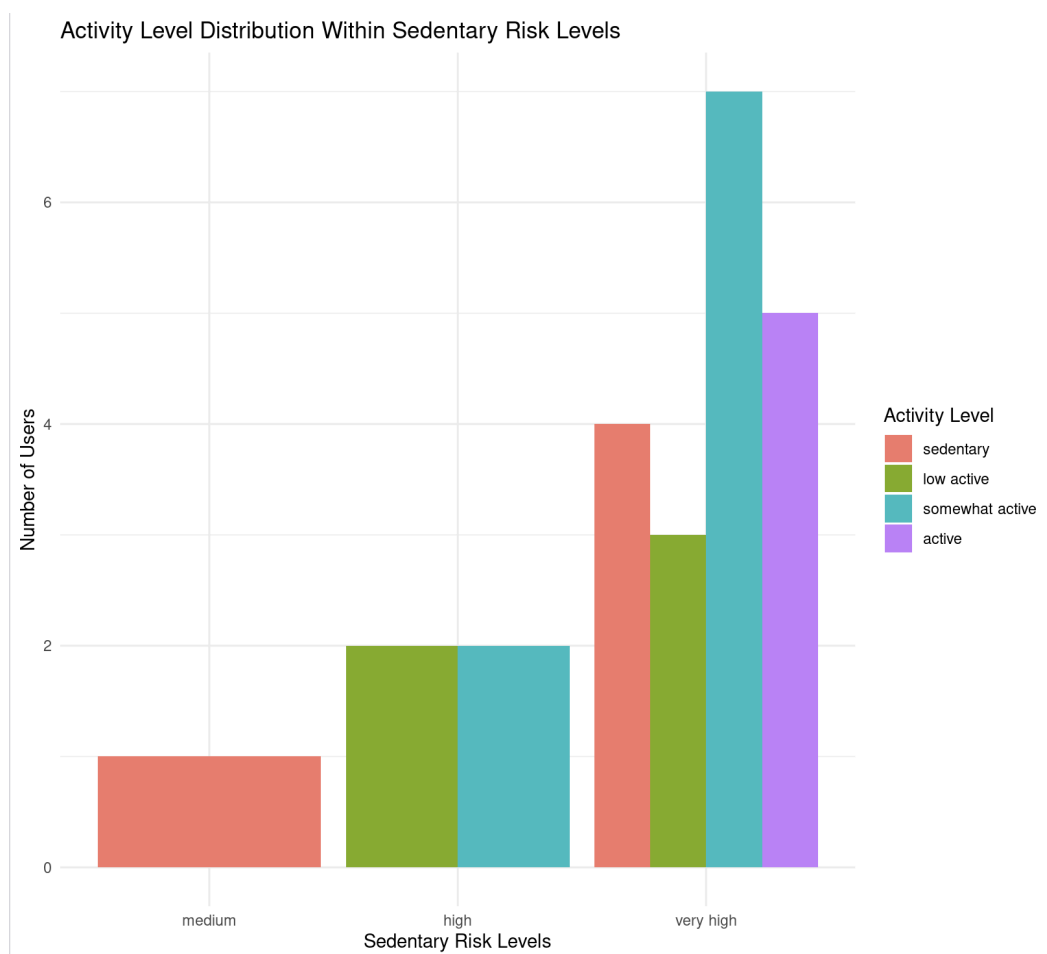
**Insights**

- ❖ Across all observations in the Activity dataset, users spent 712 minutes or about 11.9 hours each day on average in sedentary mode during the daytime and 39 minutes in bed awake not sleeping.
- ❖ The cluster of points were in the 500 daytime sedentary minutes and 30 to 50 minutes of awake time in bed.



### Distribution of Activity by Sedentary Risk Levels

Dataset: Combined Categories (Daily Activity + Sleep)



#### Insights

- ❖ A majority of users fell into the “Very High” Sedentary Risk level, spending more than 11 hours per day inactive. Within this group, users displayed a wide range of physical activity levels.
- ❖ This may indicate that some users engage in short periods of concentrated activity while still maintaining an overall sedentary lifestyle.

#### Statistical Analysis

While most variables showed weak or no significant relationships, a moderate negative correlation ( $\tau = -0.31$ ) was found between step counts and sedentary minutes from the activity data frame. This suggests that users who take more steps during the day tend to spend less time being sedentary, pointing to an inverse relationship between these two variables. Given a p-value  $< 0.05$  from the Shapiro-Wilk test, the Kendall method was considered the most appropriate for data with a small sample size and non-normal distribution.

### Step Count and Sleep Efficiency Histograms

Datasets: Activity (left) , Sleep (right)



#### Insights

- ❖ The daily step counts histogram is positively skewed, showing that most days involved moderate activity levels while a smaller number of days had significantly higher step counts. These outliers could suggest sporadic bursts of high activity across some users.
- ❖ The distribution of daily sleep efficiency is negatively skewed with a long tail toward lower efficiency. However, the majority of daily observations are clustered around 0.80, indicating most users maintained consistently high sleep efficiency.
- ❖ Together, these patterns reveal that physical activity levels vary more from day to day, while sleep efficiency tends to remain stable, with only a few instances of poor sleep quality.

## 6. Act

### Business Recommendations for Bellabeat Leaf

#### **Recommendation 1** - *Promote Consistent, Balanced Wellness Over Numeric Goals*

Insight: Many users still fall in the high risk sedentary lifestyle despite being physically active on paper.

#### Actions:

- a. Emphasize consistent and more frequent movement throughout the day
- b. Promote features like smart inactivity alerts, reminder nudges, and daily movement streaks in marketing campaigns.
- c. Highlight that the Leaf can be personalized to match each user's wellness goals, whether it's improving sleep, reducing stress, or staying gently active.

**Recommendation 2 - Highlight Sleep Tracking as a Core Feature**

Insight: Users show high sleep efficiency (mean = 0.92) but insufficient sleep duration—averaging only about 6 hours per night.

Actions:

- a. Focus on how Leaf supports healthy sleep habits through features like bedtime reminders, wind-down alerts, and smart timed morning alarms.
- b. Illustrate the product's ability to track sleep stages via the app to help users improve quality of rest, not just duration.
- c. Use holistic wellness messaging that connects sleep to overall energy, productivity, and stress management.

**Recommendation 3 - Leverage "Somewhat Active" as the Ideal Lifestyle Tier**

Insight: The "somewhat active" group had the highest proportion of users meeting recommended sleep levels, representing a realistic and balanced user profile.

Actions:

- a. Market Bellabeat Leaf as a tool for everyday wellness
- b. Less emphasis on the product as an athletic performance tool
- c. Showcase how the product can be utilized for moderate, attainable activity goals alongside sleep and stress tracking to appeal to wellness-oriented but not fitness-obsessed users.

**Future Considerations**

To strengthen this analysis, future work should incorporate more comprehensive user data, including key demographics such as gender, age, and location to better align with Bellabeat's target audience. Larger sample sizes—particularly for sleep (24 users) and weight tracking (8 users)—are needed to create more specific user segments and conduct deeper analysis. For example, understanding how BMI relates to activity levels or exploring predictive models around physical health could offer actionable insights. Additionally, richer sleep data would help uncover causes of low sleep efficiency, such as difficulty falling asleep or digital distractions, which could inform Bellabeat's marketing around rest and recovery.

### Work Cited

Barrell, Amanda . “Sitting down All Day: Consequences, Treatment, and Prevention.” [Www.medicalnewstoday.com](http://www.medicalnewstoday.com), 9 June 2023, [www.medicalnewstoday.com/articles/sitting-down-all-day#how-long-is-too-long](http://www.medicalnewstoday.com/articles/sitting-down-all-day#how-long-is-too-long).

Chaput, Jean-Philippe, et al. “Sleeping Hours: What Is the Ideal Number and How Does Age Impact This?” *Nature and Science of Sleep*, vol. 10, no. 10, 27 Nov. 2018, pp. 421–430, [www.ncbi.nlm.nih.gov/pmc/articles/PMC6267703/](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC6267703/), <https://doi.org/10.2147/nss.s163071>.

Tudor-Locke, Catrine, and David R Bassett. “How Many Steps/Day Are Enough?” *Sports Medicine*, vol. 34, no. 1, 2004, pp. 1–8, [pubmed.ncbi.nlm.nih.gov/14715035/](http://pubmed.ncbi.nlm.nih.gov/14715035/), <https://doi.org/10.2165/00007256-200434010-00001>.