Wrangling and Analyzing Insulin Clinical Trial Data

Mostafa Elseidy

Data Analyst Nanodegree, Udacity

Contents

Background 3

Introduction 4

Problem 5

Objective: fixing data issues 5

Datasets: Oral Insulin Phase II Clinical Trial Data 5

- Patients Table 5
- Treatments Table 6
- Adverse Reactions Table 6
- Additional useful information: 7

Wrangling 7

- Assessing Data 7
  - Quality Issues 7
  - Tidiness Issues 7

Analysis 8

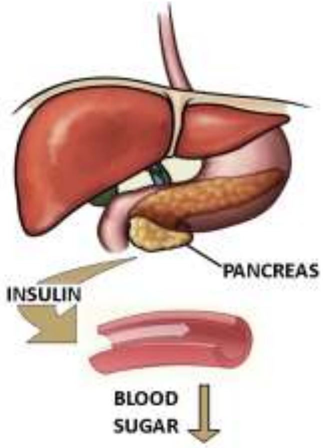
- Adverse reactions 8
- Pre-trial/Post-trial Mean Insulin Dose Change (IU) 8
- HbA1c Change 9

References 10

Background

The increasing prevalence of diabetes in the 21st century is an epidemic event. Pre 1920s, the diabetes was a feared disease that most certainly led to death. Doctors knew what it was, an uncontrollable and often elevated blood sugar levels but not how to treat it. Let alone cure it. This meant patients dealt with symptoms like unusual thirst, frequent urination, extreme fatigue and also more serious complications like stroke, blindness, loss of limbs, kidney failure and even heart attack. Luckily, in the 1920s a secretion in the pancreas that lowered blood sugar levels, soon to be called insulin, was discovered by Frederick Banting.

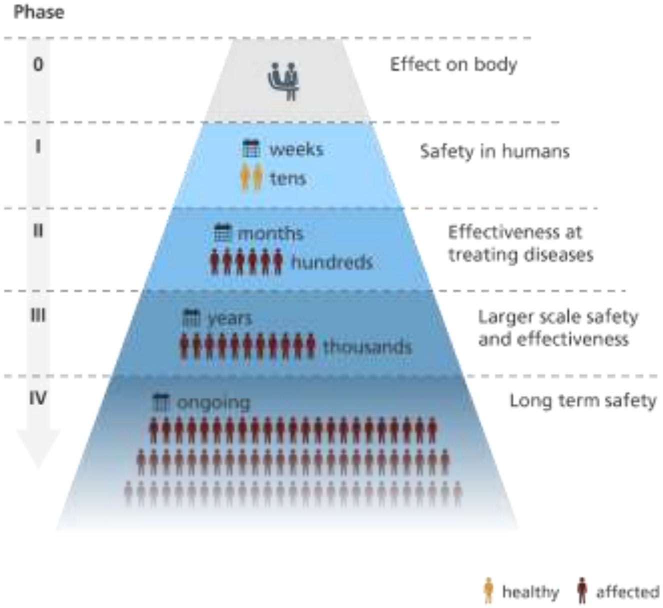
This is how insulin works, most of the food we eat is turned to glucose or sugar for our bodies to use for energy. The pancreas, an organ near the stomach, makes a hormone called insulin to help glucose get into the cells of our bodies. When you have diabetes, your body either doesn't make enough insulin or can't use its own insulin as well as it should and this causes sugars to build-up in the blood.



With Banting discovery of insulin, pharmaceutical companies began large-scale production of insulin almost immediately. Although it doesn't cure diabetes, it's one of the biggest discoveries in medicine. When it came, it was like a miracle, people with severe diabetes and only days left to live were saved. Although the default method of administration back then was a needle multiple times a day and it still is now, this is scary for some people and uncomfortable nor inconvenient for the vast majority.

Introduction

Insulin pumps are a more recent invention. These are insulin delivering devices that are semi-permanently connected to a diabetic body. But wouldn't it be great if diabetic could take insulin orally, carrying around a pocketable packet of pills, rather than a kit of vials and needles or wearing a bulky somewhat uncomfortable device? Oral insulin is the future. This is an active area of research, and has been for a long time. Historically though there's been a big roadblock, getting insulin through the stomachs thick lining.



A new innovative oral insulin drug called “Auralin”. Auralin researchers believe their proprietary capsule will solve this stomach lining problem. Phase II trials experiment test the efficacy and the dose response of the drug plus identify common short term side effects also known as “adverse reactions”. These trails typically involve several hundred patients.

In this trial, we have 350, split into two groups. One hundred seventy-five or half treated with the new oral insulin “Auralin”, and the other 175 being treated with the popular injectable insulin called “Novodra”. By comparing key metrics between these two drugs, we can determine if Auralin is effective. The most important metric, HbA1c levels, and specifically, HbA1c change.

HbA1c is a property of the blood that measures how well your blood sugar levels have been controlled, over the past few months, with higher levels being bad. If Auralin, the new oral insulin, can reduce HbA1c levels at a similar standard as the injectable insulin Novodra from some standard pretrial baseline, like say they both decrease HbA1c levels from 7.9% to 7.4%, that's a 0.5% drop. If we can get a 0.4 change, we've got ourselves a major medical breakthrough, in the dramatic quality of life improvement for diabetics all over the world. And hopefully, eventually recommend this new drug, the new oral insulin, to continue to large-scale production and improve the lives of diabetics all over the world.

Problem

Healthcare data is notorious for its errors and disorganization, and its clinical trial data is no exception. For example, human errors during the patient registration process such as having duplicate data, missing data and inaccurate data. And this common healthcare data issue is reflected in this clinical trial data. These problems block creating a trustworthy analysis.

Objective

1. Fixing data issues.
2. Analyze and visualize the clinical trial data.

Datasets: Oral Insulin Phase II Clinical Trial Data

This Auralin Phase II clinical trial dataset comes in three tables: “patients”, “treatments”, and “adverse reactions”.

Patients Table

503 patients

Columns:

- **patient_id**: the unique identifier for each patient in the Master Patient Index (i.e., patient database) of the pharmaceutical company that is producing Auralin
- **assigned_sex**: the assigned sex of each patient at birth (male or female)
- **given_name**: the given name (i.e., first name) of each patient
- **surname**: the surname (i.e., last name) of each patient
- **address**: the main address for each patient
- **city**: the corresponding city for the main address of each patient
- **state**: the corresponding state for the main address of each patient
- **zip_code**: the corresponding zip code for the main address of each patient
- **country**: the corresponding country for the main address of each patient (all United states for this clinical trial)
- **contact**: phone number and email information for each patient
- **birthdate**: the date of birth of each patient (month/day/year). The inclusion criteria for this clinical trial is age >= 18 *(there is no maximum age because diabetes is a growing problem among the elderly population)*
- **weight**: the weight of each patient in pounds (lbs)
- **height**: the height of each patient in inches (in)
- **bmi**: the Body Mass Index (BMI) of each patient. BMI is a simple calculation using a person's height and weight. The formula is BMI = kg/m^2 where kg is a person's weight in kilograms and m^2 is their height in meters squared. A BMI of 25.0 or more is overweight, while the healthy range is 18.5 to 24.9. *The inclusion criteria for this clinical trial is 16 >= BMI >= 38.*

Treatments Table

350 patients participated in this clinical trial. None of the patients were using Novodra (a popular injectable insulin) or Auralin (the oral insulin being researched) as their primary source of insulin before. All were experiencing elevated HbA1c levels.

All 350 patients were treated with Novodra to establish a baseline HbA1c level and insulin dose. After four weeks, which isn't enough time to capture all the change in HbA1c that can be attributed by the switch to Auralin or Novodra:

- 175 patients switched to Auralin for 24 weeks
- 175 patients continued using Novodra for 24 weeks

Columns:

- **given_name**: the given name of each patient in the Master Patient Index that took part in the clinical trial
- **surname**: the surname of each patient in the Master Patient Index that took part in the clinical trial
- **auralin**: the baseline median daily dose of insulin from the week prior to switching to Auralin (the number before the dash) *and* the ending median daily dose of insulin at the end of the 24 weeks of treatment measured over the 24th week of treatment (the number after the dash). Both are measured in units (shortform 'u'), which is the international unit of measurement and the standard measurement for insulin.
- **novodra**: same as above, except for patients that continued treatment with Novodra
- **hba1c_start**: the patient's HbA1c level at the beginning of the first week of treatment. HbA1c stands for Hemoglobin A1c. The HbA1c test measures what the average blood sugar has been over the past three months. It is thus a powerful way to get an overall sense of how well diabetes has been controlled. Everyone with diabetes should have this test 2 to 4 times per year. Measured in %.
- **hba1c_end**: the patient's HbA1c level at the end of the last week of treatment
- **hba1c_change**: the change in the patient's HbA1c level from the start of treatment to the end, i.e., hba1c_start - hba1c_end. For Auralin to be deemed effective, it must be "noninferior" to Novodra, the current standard for insulin. This "noninferiority" is statistically defined as the upper bound of the 95% confidence interval being less than 0.4 for the difference between the mean HbA1c changes for Novodra and Auralin (i.e. Novodra minus Auralin).

Adverse Reactions Table

Columns:

- **given_name**: the given name of each patient in the Master Patient Index that took part in the clinical trial and had an adverse reaction (includes both patients treated Auralin and Novodra)
- **surname**: the surname of each patient in the Master Patient Index that took part in the clinical trial and had an adverse reaction (includes both patients treated Auralin and Novodra)

- **adverse_reaction**: the adverse reaction reported by the patient

Additional useful information:

- Insulin resistance varies person to person, which is why both starting median daily dose and ending median daily dose are required, i.e., to calculate change in dose.
- It is important to test drugs and medical products in the people they are meant to help. People of different age, race, sex, and ethnic group must be included in clinical trials. This diversity is reflected in the "patients" table.

Wrangling

Assessing Data

Quality Issues

Patients Table

- Zip code is a float not a string
- Zip code has 4 digits sometimes
- Tim Neudorf entry height is 27 in instead of 72 in
- Full state names sometimes, abbreviations other times
- Typo in the given name column for Dsvid Gustafsson. Dsvid instead of David
- Missing demographic information (address – contact columns)
- Erroneous datatypes (assigned sex, stat, zip code, and birthdate columns)
- Multiple phone number formats
- Default John Doe data (duplicated 6 times)
- Multiple records for Jakobsen, Gersten, Taylor
- Kgs instead of lbs for Zaitseva weight

Treatments Table

- Missing HbA1C changes
- The letter "u" in starting and ending doses for Auralin and Novodra columns
- Lowercase given names and surnames
- Missing records (280 instead of 350)
- Erroneous datatypes (auralin and novodra columns)
- Inaccurate HbA1C change (4s mistaken as 9s)
- Nulls represented as dashes (-) in auralin and novodra columns

Adverse Reactions Table

- Lowercase given names and surnames

Tidiness Issues

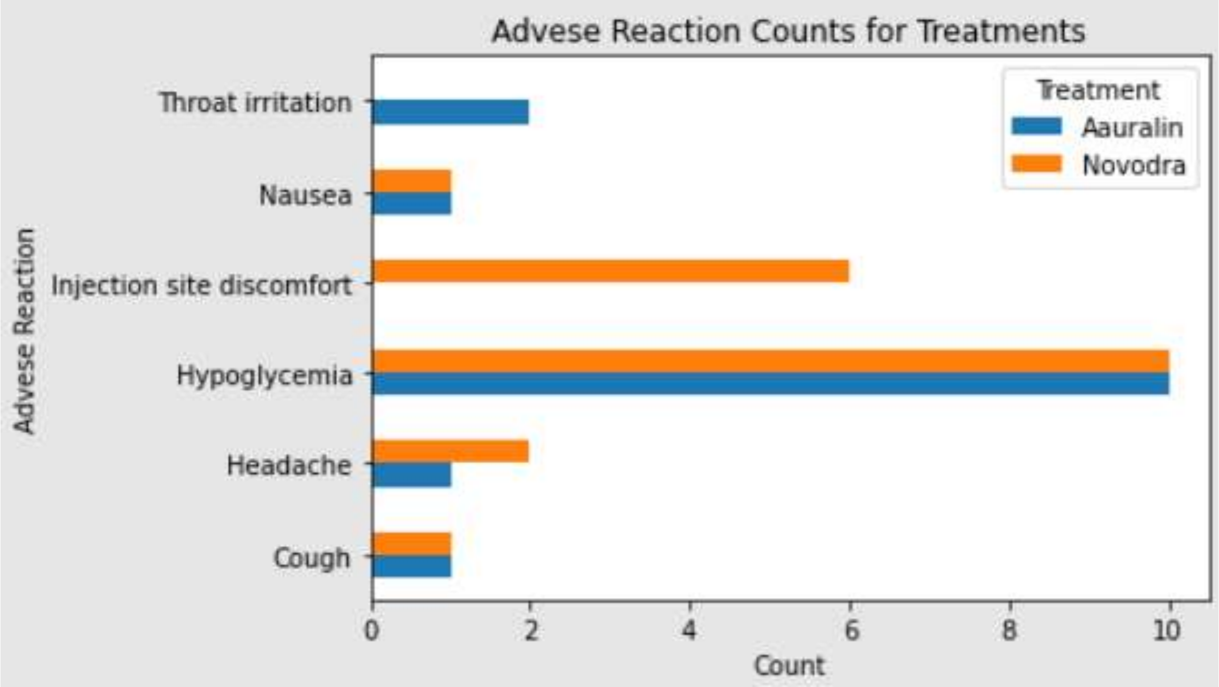
- Contact column in "patients" table should be split into phone number and email
- Three variables in two columns in "treatments" table (treatment, start dose and end dose)

- Adverse reaction should be part of the “treatments” table
- Given name and surname columns in “patients” table duplicated in treatments and “adverse_reactions” tables

Analysis

To pass phase II, Auralin must be safe and effective. to be deemed effective, it must be "noninferior" to Novodra.

Adverse reactions



The side effects are similar for Auralin and Novodra except for:

1. Hypoglycemia, which is low blood sugar caused by an insulin overdose, is the most adverse reaction
2. Throat irritation is a side effect of Auralin, which can be expected because this bill is taken orally and passes by the throat before it gets to the stomach.

Pre-trial/Post-trial Mean Insulin Dose Change (IU)

If Auralin requires a way higher dosage to be effective, the increase in cost will affect the profit and the manufacturer might not bring that to the market. From business or industry perspective, this is one of the essential concerns for the manufacturer to make things financially feasible.

Comparing dose change means, the patients that were treated with Auralin required on average 8 more units of insulin to establish a safe steady blood sugar level. For Novodra, patients on average required about 0.4 units less of insulin.

The fact that Auralin required more units is kind of expected because we knew that oral insulin had a tougher time in getting to the bloodstream through the stomach lining. So, the results are good for Auralin or at least they are not bad.

HbA1c Change

This is a key indicator for diabetes control. For Auralin to be deemed effective, it must be noninferior to Novodra, the current standard for insulin. This “noninferiority" is statistically defined as the upper bound of the 95% confidence interval being less than 0.4 for the difference between the mean HbA1c changes for Novodra and Auralin (i.e., Novodra minus Auralin).

Coeff.	Std. err	t	P>|t|	[0.025	0.975]
0.0171	0.006	2.710	0.007	0.005	0.029

A mean of 0.4 for Novodra and 0.39 for Auralin. After conducting statistical analyses, we reject the null hypothesis in favor of alternative one. It is more likely that there is no significant difference for HbA1c change between Novodra and Auralin. As a result, (Auralin) oral insulin is similarly effective to (Novodra) injectable one and it does not break the upper limit. According to confidence interval of difference (0.0047, 0.0294), the upper limit (0.03) is less than 0.4.

References

https://mstranslate.com.au/understanding-clinical-trial-process/

https://healthjade.net/hba1c/

http://media.hypersites.com/clients/1446/filemanager/Articles/DocCenter_Problem_with_data.pdf