



## Dataset: Oral Insulin Phase II Clinical Trial Data

---

### DISCLAIMER: This Data Isn't "Real"

The Auralin and Novodra are **not** real insulin products. This clinical trial data was fabricated for the sake of this course. When assessing this data, the issues that you'll detect (and later clean) are meant to simulate real-world data quality and tidiness issues.

That said:



- This dataset was constructed with the consultation of real doctors to ensure plausibility.
- This clinical trial data for an alternative insulin was inspired and closely mimics this real [clinical trial for a new inhaled insulin called Afrezza](#).
- The data quality issues in this dataset mimic real, [common data quality issues in healthcare data](#). These issues impact quality of care, patient registration, and revenue.
- The patients in this dataset were created using this [fake name generator](#) and do not include real names, addresses, phone numbers, emails, etc.

The video above is only a short preview of the dataset that is intended to motivate. So don't worry if the details don't all make sense right now. You'll get intimately familiar with each column in each table in the dataset shortly. If you want to dive deeper into the data now, hop ahead to the **Visual Assessment: Acquaint Yourself** page where the data files are provided in a Jupyter Notebook workspace. (You can also download the files from there if you'd like by clicking the Jupyter logo in the workspace then selecting and downloading each file.)

Have questions? Head to the [forums](#) for discussion with the Udacity Community.

NEXT