

SEARCH



RESOURCES

CONCEPTS

- ✓ 3. Dataset: Oral Insulin Phase II C...
- ✓ 4. Unclean Data: Dirty vs. Messy 1
- 5. Unclean Data: Dirty vs. Messy 2
- 6. Assessment: Types vs. Steps
- 7. Visual Assessment
- 8. Visual Assessment: Acquaint Yo...
- 9. Quality: Visual Assessment 1
- 10. Assessing vs. Exploring
- 11. Quality: Visual Assessment 2
- 12. Data Quality Dimensions 1
- 13. Data Quality Dimensions 2



Mentor Help

Ask a mentor on our Q&A platform



Peer Chat 3

Chat with peers and alumni



Int J Health Sci (Qassim)

Int J Health Sci (Qassim). 2007 Jul; 1(2): V–VIII.

PMCI

Is Diabetes Becoming the Biggest Epidemic of the Twenty-first Century?

Syed Amin Tabish, FRCP, FRCPE, FAMS, FACP, Editor-in-Chief[Author information](#) ► [Copyright and License information](#) ►This article has been [cited by](#) other articles in PMC.

Diabetes is a major public health problem that is approaching epidemic proportions globally. Worldwide, the prevalence of chronic, noncommunicable diseases is increasing at an alarming rate. About 18 million people die every year from cardiovascular disease, for which diabetes and hypertension are major predisposing factors. Today, more than 1.7 billion adults are overweight, and 312 million of them are obese. In addition, at least 155 million children worldwide are overweight or obese. A diabetes epidemic is underway. According to an estimate,

DISCLAIMER: This Data Isn't "Real"

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

That said:

- This dataset was constructed with the consultation of real doctors to ensure accuracy.
- This clinical trial data for an alternative insulin was inspired and closely modeled after a [clinical trial for a new inhaled insulin called Afrezza](#).
- The data quality issues in this dataset mimic real, [common data quality issues](#). These issues impact quality of care, patient registration, and revenue.
- The patients in this dataset were created using this [fake name generator](#). We used their real names, addresses, phone numbers, emails, etc.

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