**Part 1: Validate and Verify Data and Grading Criteria**

**Step 1: Verification of Data**

*In what ways was the dataset reviewed to ensure it remains representative of the Medicare population it is intended to match?*

De-SynPUF was synthesized from actual datasets rather than presenting the real datasets as to avoid exposing patient identifiers. The trends, counts, and aggregates were maintained in the synthetic generation though only accounts for 5% of the original 2008 Medicare Beneficiary dataset. The following below are the six points of De-SynPUF datasets

1. **Number of Claims per Beneficiary by Service Type Over Three Years**
2. **Demography**
3. **Claimants by Service Type by Year**
4. **Numbers of Claims of Claimants by Service Type by Year**
5. **Reimbursement by Source by Year**
6. **Various chronic conditions**

*Describe at least four data elements which were reviewed.*

On demographic level:

1. Sex
2. Race
3. Year of Birth

These Medicare data elements were synthetically imitated to be as close to the Census dataset. High reliability and data quality can be assumed for the demographics for this synthesize dataset.

On Condition level:

1. Chronic condition in 2008

Comparing this data element from the actual dataset, it is worth noting that in general, the chronic disease prevalence a discrepancy of more than 25% can be seen. Thus, the wide margin would produce a questionable and varying results and replication of the study and reproducibility of the result would unattainable.

*What identifiers are used to enable linking of patient records across systems?*

DESYNPUF\_ID - unique identifier for the beneficiary code

CLM\_ID - unique identifier for the claim

**Step 2: Validation of Clinical Data**

Find the distribution of ages for the population in this database and recent census data describing the age distribution of the overall US Population and answer the following questions:

*How does the Medicare data set compare with the overall US population?*

The Medicare dataset have more senior population or technically it is skewed to the right. The age distribution focuses on the 70-79 year-old which comprises to 60% of the dataset distribution. On the contrary, the Census data aging 65 years and above accounts only to 16% of the total population.

*What parts of the population are represented in the Medicare data set?*

The representation of beneficiaries are mostly on white elder population aging 77 years and above.

*What parts are not represented in the Medicare data set?*

There is a noticeable underrepresentation of younger populace (i.e. below 77 years old) which is about 70% of the total population according to the Census dataset. Moreover, non-white population were also not faithfully represented.

*Describe analyses where this data set provides sufficient quality to support its use, kinds of analysis where it would not be appropriate.*

The datasets may provide sufficient quality if the analysis of claims are mostly focused on the elderly white population whose conditions or disease are similar to those represented in the data (e.g. Alzheimers, Diabetes, etc.. ). However, it would be inappropriate to use the same dataset if the results will be use to infer to the greater population; thus, inferential analysis and statistics is not deemed useful for this use case.

**Part 2: Why Doctors Hate Their Computers**

**Gregg Meyer**. The statement of Gregg Meyer that *“We think of this as a system for us* (i.e. clinicians) *and it’s not. It is for the patients”* debatable. However, I totally agree with him knowing that at the end of the day, the patient’s welfare must outweigh everything – as this is the oath every medical practitioner took and signed up for. In summary, two main contradicting ideas were raise: (1) most medical professionals may believe that the inefficiencies of these systems is a great contributor to their burn-outs – thus leading to less people caring for the patients and (2) these systems, however clunky they are – are making real improvements in care. Meyer pointed out the benefit of using these health information systems (HIS) at some point outweighs the annoyance to the main end-users (i.e. health workers). Computerization allows clinicians to help patients in ways that had not been possible before. And this is demonstrably true in a lot of cases. Concretely, this was shown in one of the patient case where an insidious parathyroid tumor was diagnosed by digging deeper into the patient records – which would have been impossible if it was done manually. I also concur that we are just only in the beginning – we will encounter a lot of hurdles along the way but in the long term, we will get more benefit as the software will evolve and make things less complicated.

**Neil Malhotra.** Among all the mentioned personalities, it was Neil Malhotra that actually did more than complain – he tinkered and found a way to solve the inefficiencies he encountered. The neurosurgeon devised a plan by convening *“an open weekly meeting, currently on Thursday mornings, where everyone in the neurosurgery department—from the desk clerks to the medical staff to the bosses—could come not just to complain about the system but also to reimagine it”.* Similar to everyone, it was inevitable for him to experience displeasure to these systems. But unlike the rest, he tried to find a solution and successfully overcame the resistance that goes with it. He gathered the concerned stakeholders and received the feedback for constructive resolution of the problem. In the end, their department came with a better system that captures only the most relevant information without being burdensome. The results of this simple activity came into fruition through a combination of his ingenuity and group collaboration. Like in any problem we encounter in life, we must not dwell on the hurts and pains it cause – but rather find a way to resolve it. We cannot just keep on giving desultory complaints and be a Luddite curmudgeon. We must have the resilience to prevail over the problem, but more than that it is mandatory that we find a solution towards it. These so called annoyance, inefficiencies, and pointless systems are valid problems and concerns; thus, like Malhotra we can always find a way though it.

**David Woods.** As a systems theorist, it is but expected for David Woods to centered his ideas around systems rather than humans. Stating that *“it’s the people inside organizations, not the machines, who must improvise in the face of unanticipated events”* is a valid premise from his own perspective - however, I beg to differ. Understandably, it was just recently that the era of user experience and design thinking begun to flourish. Prior to this period, legacy enterprise software ruled the workplace more than ever – the user process and workflows were change just to accommodate the enterprise systems. However, things have change and people are now place at the center of the experience. User experience design and stakeholders’ consultation using design thinking workshops were already part of the norm prior to system development and implementation. The statement to *“install a monolith, and the smallest changes require a committee decision, plus weeks of testing and debugging to make sure that fixing the daylight-saving-time problem, say, doesn’t wreck some other, distant part of the system”* at some point have a degree of mendacity. Regression testing, version control, continuous integration and deployment are just some of the progress of software lifecycle development to address this valid concerns. The truth of tomorrow may not be the truth of today. Concretely, it was true that in the past that monolith legacy system entails and ginormous effort for a slight feature changes. Today, this can be done seamlessly without a downtime notice.

For the chosen viewpoints, these are the solutions to overcome the requirements: (1) have rigorous stakeholder consultation for the end-user requirements, (2) design the enterprise system to be highly customizable accordingly, (3) give importance to user experience, (4) have a design thinking methodology to approach the problem, (5) opt for open interoperability by having accessible API endpoints, and lastly (6) focus on the patient more than ever. As an effective health data analyst, I need to be aware of the following concepts: (1) form and signal fatigue are real and not just a slight annoyance of clinicians, (2) revenge of the ancillaries are caused by poor data and form designs, (3) basket issue in problem list are result of poor user experience and graphical interface designs. These perspectives affect mostly the clinicians as they are contributing factor towards compassionate fatigue and burn-out. In general, bigger data will lead to bigger outcomes. Though also true that burn-out will contribute to shortage of physicians, there are always a midpoint were we can find a common ground to solve the issue. Machine Learning and artificial intelligence can be seen as the golden opportunity to address the issue. Bigger data will definitely benefit the training of model in an AI-assisted health systems. Smarter and more intelligent HIS will thus remove the heavy burden that current physicians are experiencing. So it is a win-win situation, as more data are gathered will lead to smarter HIS and definitely to less burden to physicians.