Analysis of Hospital Pricing Data

ABIGAIL SLATER, DESTINY ZIEBOL, TYSON REED, and GAVIN SCHAEFERLE

This report develops a model that can be used to standardize publicly available inpatient Hospital Price Transparency data.

1 INTRODUCTION

Transparency in healthcare pricing is a component of ever evolving healthcare legislation in the U.S. As new pricing publication guidelines are implemented, exciting new opportunities for analysis and patient support become available. As the Centers for Medicare and Medicaid Services (CMS) Hospital Price Transparency Rule went into effect in 2021, a patchwork of pricing data has become publicly accessible. Group 11 developed this project aiming to collect a selection of this publicly accessible data and integrate it into a single database. The intentions of this database were to provide a platform for comparison and analysis between hospitals and hospital systems in addition to exploring possible warehousing needs for integrating this highly varied hospital data.

2 BACKGROUND

The 2010 U.S. Affordable Care Act required American hospitals to publish a list of standard charges for services and items provided at every hospital location. After eight years of opposition, hospitals began publishing chargmaster rates in 2018, at which time it became clear that unnegotiated pricing data did not reflect the charges seen by most patents. In June of 2019, president Donald Trump issued an executive order requiring the development of standards for publishing negotiated rates. The U.S. Health and Human Services department (HHS), along side CMS developed standards and timelines for the publication of these data, with final publishing guidance released in November of 2019. Current guidelines went into effect January 1, 2021. In July of 2021, President Joe Biden farther endorsed the release of negotiated hospital pricing data with an executive order supporting the current publishing rules and directing continued transparency despite push back from the healthcare industry, which largely viewed such data to be a trade secret. [HollandKnightHealthcare 2021]

Hospitals in the United States are now required by law to publish their negotiated rates with Managed Care Organizations (MCO) online in a machine-readable format. The intent of the law is to provide consumers with price transparency that will permit Americans to shop and make informed choices about the cost of their care. However, there is limited guidance from the Federal Government on how hospitals are to comply with the law (e.g. no standardization of file formats, lack of standardized naming convention, etc.)

As of today, most hospitals have published their rates with MCOs on their website and are in compliance with the law, but the approach used by each hospital is different. Some hospitals published rates in a comma separated value (CSV) format, a text file, or some other flat file format. Other hospitals require users to input various items on a website (E.g. the service and insurance coverage information) and then return the rates specific to that individual [Northfield 2021]. Given the multitude of permissible approaches and data locations, Group 11 decided to aggregate information into one central location to permit easier comparison and analysis.

The use-case of this model is broad. While the intent was to allow individuals clearer price transparency, there are other stakeholders that could use this information. Insurance companies can use it to better understand new markets they are planning to enter. Also, hospital systems (or insurance companies) can use it to gain negotiation leverage (e.g. you are paying our competitor 10% more for the same service, you need to increase our rates). Similarly, governmental authorities can use it to help regulate and understand hospital payment trends.

For this project, Group 11 aggregated three hospital systems operating in Minnesota (Essentia, Mayo, and Fairview) into one database structure. This approach provides a template that could be scaled and applied to all hospital systems operating in the United States. This report outlines Group 11's approach, the overall data architecture, and challenges encountered while aggregating hospital pricing data. This project is not intended to be a comprehensive review of the law or a review of the selected hospital's compliance with regulations. Instead, this project is an outline of how hospital pricing data can be structured and standardized to improve access to knowledge of hospital prices. The tool developed in this report is not a front-facing user program, but a data warehouse that could potentially back a front-facing user program.

3 MODEL DEVELOPMENT

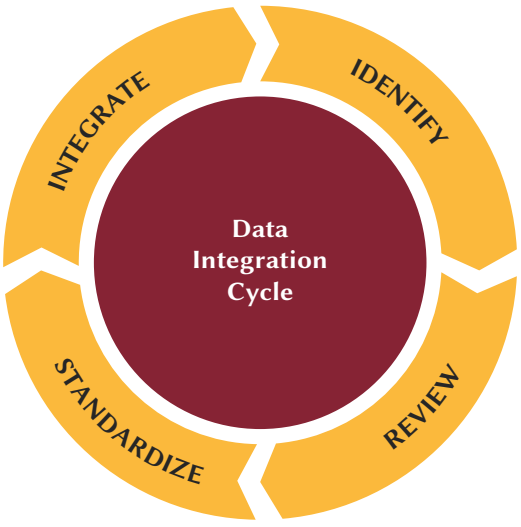

Fig. 1. Steps used to integrate each hospital into database.

3.1 Database Hosting

This database was hosted though Azure SQL Database services. Azure was chosen due to its high flexibility and market presence.

Should this project expand, Azure database hosting enables rapid expansion and scalable support for font end traffic. Queries and uploads were executed though Microsoft SQL Server Management Studio.

3.2 New Data Integration

Group 11 used the cycle shown in Figure 1 when integrating a new hospital into the database. For each new hospital provider, the group **Identified** the pricing data on the hospital's website. Then, the team **Reviewed** the data structure underlying the hospital's pricing data. Once the review was complete, the team **Standardized** the hospital's pricing data (e.g. mapped to standard naming convention). After this step, the team **Integrated** the hospital's pricing data into the database.

As hospitals update pricing data (e.g. after contracts with MCOs change or on January 1st of each year), this process would be repeated and could be automated. Automating the data export, load, and transforming component of this project would require process checks to ensure that hospitals did not make changes to the data structure or formats. Since each hospital's data structure is unique, developing and maintaining these additional checks would be expensive and time consuming.

3.3 Data Security

The hospital pricing transparency data intended for this database is based entirely on publicly available data. As such, the security required to protect this data from malicious activity is minimal. However, the structure of this database could be considered proprietary and access to working tables should be limited to individuals that understand and know how to use it.

Data consistency is the responsibility of the user, and given the extensive cleaning and pre-processing needed to prepare data for database integration, this responsibility should not be taken lightly. This database does not currently employ check constraints or views for controlling access, although table constraints were added as much as possible to ensure proper use.

3.4 Transaction Management

Updating this database will take place rarely (e.g. once a year when contracts between hospitals and MCOs are updated). As such, there will be very few circumstances where write queries are used in the database. Said differently, the possibility of *Reading Uncommitted Data, Unrepeatable Reads, and Overwriting Uncommitted Data* conflicts in this database is small and can be controlled by temporarily locking the entire database while updates occur.

4 DATA CHALLENGES

4.1 Inconsistent Naming Convention

Each hospital network used different naming conventions for the MCOs and Providers, which were standardized before adding to the database. For example, 'BCBS' was standardized to Blue Cross Blue Shield.

4.2 Inconsistent Data

It was found that not all providers published rates for all services. For example, some providers used attribute values such as *no volume* to denote that there wasn't enough hospital admissions within the MS-DRG code to calculate a credible average charge. Other providers published *n/a* in this scenario. One hospital (CentraCare Long-Prairie) had error values (#VALUE!) that were not scrubbed prior to publishing rates. These occurred in the *Minimum* and *Maximum* columns, so this does not represent an issue with publishing prices, but illustrates the difficulty in managing streamlining loading and transforming data. Some hospitals included a row for each unique MCO/Product combination, but only included cost for some which created sparse tables with many NULL values. It was decided to remove these NULL values, as well as Per Diem and average values.

4.3 Data Format

Each hospital system examined published their data in different formats. For example, some hospitals used a long or wide format on a single sheet to differentiate rates between MCOs and Products while other hospitals used multiple excel tabs to differentiate between rates. Formatting was standardized before integration into the database.

4.4 Data Location

There is no centralized location for hospital price data, either on a state or federal level. Each hospital systems published rates are located on their corresponding websites. Unfortunately we found that the data was not in a similar location from website to website, as some hospital systems included a centralized page for each of their locations pricing data whereas other hospitals included the pricing data on each of the specific locations main websites.

4.5 Data Accessibility

Some hospitals provided published rates in a machine readable file and others provided a website tool. For example, Northfield Hospital has a website tool that permits the user to select a service category (e.g. *Bone Scan*), and a service (e.g. *Bone Density Scan*) [Northfield 2021]. The user is then required to submit patient information (e.g. name, date of birth, gender, etc.) and insurance information (e.g. insurance company, group number, member identifier, etc.). The tool then provides a cost estimate to the customer. Only those hospitals that published machine readable files were included in our data set.

4.6 Data Interpretation

Interpreting pricing data is difficult given the structure of how hospitals are reimbursed for providing services. Hospitals maintain a fee schedule referred to as the *Charge Master* which is what a hospital will bill to an MCO for an admission. In most circumstances, the MCO will have a negotiated rate for the hospital (e.g. 105% of the Medicare fee schedule, 80% of billed charges, $1,000 per diem, etc.) that is different than the billed amount. This is the total amount that is actually paid to the hospital for the service and is referred to as the *allowed amount*. The *allowed amount* is paid by MCO in addition to the member through deductibles, coinsurance, or copyaments.

The amount the member pays is referred to as the *member liability* and the amount the MCO pays is the *plan liability* or *net paid*.

It isn't clear which rates are published (e.g. the *charge master*, *allowed amount*, etc.). Furthermore, the amount that is actually paid can vary by admission even within the same DRG code. For example, the Medicare Inpatient Prospective Payment System (IPPS) includes an outlier provision and new technology add-on payments for certain types of admissions. Given these complexities, its difficult to interpret the data published by the hospitals.

5 DATA MODELS

5.1 Entity Relationship (ER) Diagram

Group 11 developed the ER diagram shown in Figure 2 as a simplified provider contracting model within an insurance market context. From a high-level, each MCO has a unique identifier (*mid*) and a MCO name (*mnane*). Each MCO can, but is not required to, offer products to its customers. Each product must have a single corresponding MCO, a unique identifier (*pid*) and a product name (*pname*). The product covers a set of services that are identified by a diagnostic related group code (*drg_cd*). Each *drg_cd* has a description (*drg_desc*) that provides a description of the service. For example, the *drg_cd* = 001 is coded for a heart transplant without major complication or comorbidities. Each hospital is uniquely identified through an attribute labeled *hid*, has a name referred to as the *hname* and provides services (e.g. a heart transplant) identified by the *drg_cd*. The cost of each service provided by a hospital to a member enrolled in a specific product is referred to as the *allowed* amount. The total allowed amount for an admission includes both the member cost sharing (e.g. deductible, copay, coinsurance, etc.) and the amount the MCO paid for the admission (e.g. plan paid amount). The split between member cost-sharing and plan cost-sharing is not included in the ER diagram due to the complexities of adjudicating claims.

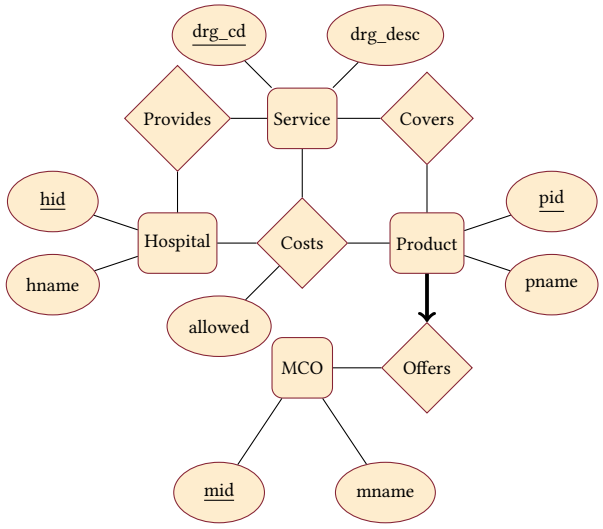

Fig. 2. Hospital Pricing Transparency Entity Relationship Diagram

5.2 Relational Schema and Table Descriptions

The tables below translate the ER diagram into a schema that provides data types, and are followed by a description of each schema and relation.

Hospital(*hid*: **integer**, *hname*: **string**)
Service(*drg_cd*: **string**, *drg_desc*: **string**)
Product(*pid*: **integer**, *pname*: **string**)
MCO(*mid*: **integer**, *mname*: **string**)
Provides(*hid*: **integer**, *drg_cd*: **string**)
Covers(*pid*: **integer**, *drg_cd*: **string**)
Offers(*pid*: **integer**, *mid*: **integer**)
Costs(*drg_cd*: **string**, *hid*: **integer**, *pid*: **integer**, *allowed*: real)

Hospital - This table maps an identifier (*hid*) to a hospital name. CMS Certification Number (CCN) was used as as the *hid* value for a given hospital for consistency. [CMS 2021].

Service - This table maps the service identifier to the description. The identifier is the 3-digit Medicare Severity-Diagnosis Related Group (MS-DRG) and the description was sourced from publicly available mappings [ICD 2022]. To ensure consistency, this relation includes a check that the length of the DRG code equals three. DRG codes are numeric values between 1 and 999 with leading zeros (e.g. 001). Leading zeros are a part of the coding system shared nation wide, and as numeric data types do not preserve leading zeros this field was stored as characters. The *drg_desc* is a brief description of the *drg_cd*.[1]

Product - This table maps an identifier (*pid*) to a product name. These identifiers were given to the product names alphabetically (0-63), thus are unique to our database.

MCO - This table maps an identifier to the name of the Managed Care Organization (MCO). This table was developed from the hospital transparency data and standardized across providers. For example, if one hospital system referred to a MCO as *Blue Cross Blue Shield of Minnesota* and another *BCBSMN*, then the team used standardized naming conventions to ensure consistency between hospitals.

Provides - This relation includes a list of DRG codes for which a hospital has published rates. The primary key is *hid* and *drg_cd*. *hid* is a foreign key that references Hospital and *drg_cd* is a foreign key that references Service.

Covers - This relation provides the list of *pid* values from the product table and *drg_cd* values from the Services table. The primary key of this table is *pid* and *drg_cd*. *pid* is a foreign key that references product and *drg_cd* is a foreign key that references Service. These foreign key constraints ensure that each combination of *pid* and *drg_cd* have a meaningful interpretation.

[1]For more information regarding DRG codes, please see https://en.wikipedia.org/wiki/Diagnosis-related_group.

Offers - This relation provides the list of *pid* values from the product table and *mid* values from the MCO table. The primary key of this table is *pid* and *mid*. *pid* is a foreign key that references product and *mid* is a foreign key that references MCO. This table provides unique combinations of MCO and Products, displaying which products are offered by each MCO.

Costs - This relation provides the published cost (*allowed*) for each given hospital, service and product combination. The primary key for this table is the *hid*, *pid*, and *drg_cd*. This table has a domain constraint that *allowed* must be numeric (e.g. *no volume* and *n/a* entries are not permitted) and an integrity constraint that *allowed* must be greater than or equal to $0.00.[2] Said differently, a hospital should not pay an MCO to provide the service. *hid* is a foreign key and references Hospital, *pid* is a foreign key and references Product, and *drg_cd* is a foreign key and references Service.

5.3 Data Definition Language (DDL)

Below outlines the SQL create statements for each of the tables defined in the schema at the start of section 5.2.

```
CREATE TABLE Hospital (hid INTEGER,
                       hname CHAR(20),
                       PRIMARY KEY (hid))

CREATE TABLE Service (drg_cd CHAR(3),
                      drg_desc CHAR(200),
                      PRIMARY KEY (drg_cd),
                      CHECK (LEN(drg_cd) = 3))

CREATE TABLE Product (pid INTEGER,
                      pname CHAR(200),
                      PRIMARY KEY (pid))

CREATE TABLE MCO (mid INTEGER,
                  mname CHAR(200),
                  PRIMARY KEY (mid))

CREATE TABLE Provides (hid INTEGER,
                       drg_cd CHAR(3),
                       PRIMARY KEY (hid, drg_cd),
                       FOREIGN KEY (hid)
                            REFERENCES Hospital,
                       FOREIGN KEY (drg_cd)
                            REFERENCES Service)

CREATE TABLE Covers (pid INTEGER,
                     drg_cd CHAR(3),
                     PRIMARY KEY (pid, drg_cd),
                     FOREIGN KEY (pid)
                        REFERENCES Product,
                     FOREIGN KEY (drg_cd)
                        REFERENCES Service)

CREATE TABLE Offers (mid INTEGER,
                     pid INTEGER,
                     PRIMARY KEY (mid, pid),
                     FOREIGN KEY (mid)
                        REFERENCES MCO,
                     FOREIGN KEY (pid)
                        REFERENCES Product)

CREATE TABLE Costs (hid INTEGER,
                    drg_cd CHAR(3),
                    pid INTEGER,
                    allowed REAL,
                    PRIMARY KEY
                        (pid, drg_cd, hid),
                    FOREIGN KEY (pid)
                        REFERENCES Product,
                    FOREIGN KEY (drg_cd)
                        REFERENCES Service,
                    FOREIGN KEY (hid)
                        REFERENCES Hospital,
                    CHECK (allowed >= 0))
```

[2] In actual claim data, negative allowed amounts are permitted for denied or voided claims that offset original or replaced claims.

5.4 Indexes

The following indexes were used on the database to improve processing speed and data organization. Each of the datasets are relatively small with the exception of *Costs*. As such, the time to maintain indexes while updating the tables is minimal.

(1) Hospital - Clustered index on *hid* and non-clustered index on *hname*.
(2) Service - Clustered index on *drg_cd* and non-clustered index on *drg_desc*.
(3) Product - Clustered index on *pid* and non-clustered index on *pname*.
(4) MCO - Clustered index on *mid* and non-clustered index on *mname*.
(5) Provides - Clustered index on *hid, drg_cd*.
(6) Covers - Clustered index on *pid, drg_cd*.
(7) Offers - Clustered index on *mid, pid*.
(8) Costs - Clustered index on *allowed* and non-clustered index on each of *hid*, *drg_cd* and *pid*. The use of a clustered index on *allowed* permits ranged searches (e.g. What are the services that cost between $15,000 and $25,000?).

5.5 Table Population

After each of the hospital systems raw data was cleaned and standardized, the three datasets were uploaded to their own respective tables in the database. These tables were then merged into one "basedata" table that included six standard fields–hospital, MCO, product, DRG code, DRG description and allowed cost–for all three hospital systems. This basedata table was used in SQL queries to pull the necessary data for each of the tables created in section 5.3. Using

the 'provides' relation as an example, the following query was used to insert data:

```
INSERT INTO Provides
select distinct H.[hid],
                A.[drg_cd]
FROM [BaseData] A
left join [Hospital] H
on A.[Hospital] = H.[hname]
```

This query joined the hospital table (which includes the unique hospital IDs) with the basedata table to obtain the unique combinations of hospital IDs and DRG codes. This approach of selecting distinct combinations from the basedata table and joining to other tables already created was used to populate each of the eight tables in the database.

6 RESULTS AND EXAMPLE QUERIES

Once the insert queries were complete, we could started pulling read queries from our data base as analysis results. The results from this database can be split into three parts–basic pulling of data in the form of sql queries, comparing hospitals and providers by calculating z-scores per DRG code, and implementing a simple Rshiny dashboard for use of looking up information from our data base.

6.1 Result One

The first result of this database consists of pulling basic hospital pricing information. Such information includes the allowed amount for a given Hospital, MCO, and Product.

Example One: Identify the top 100 products with the lowest allowed amount for any hospital for the given DRG code 177.

```
select top 100 hos.hname,
           mco.mname,
           pro.pname,
           min(cost.allowed) as minCharge
from dbo.Costs cost
inner join dbo.Hospital as hos on hos.hid = cost.hid
inner join dbo.Product prod on prod.pid = cost.pid
inner join dbo.Offers as ers on prod.pid = ers.pid
inner join dbo.MCO as mco on ers.mid = mco.mid
where cost.drg_cd = '177'
group by hos.hname,mco.mname,pro.pname
order by minCharge
```

This query returned a table of the lowest allowed amount for a given hospital, MCO, and product. Given this specific example of DRG code 177 (an "influenza like Respiratory illness with Major Complications") the lowest allowed amount was for Essentia Fosston with Health Partners product "CARE MSHO" at the price of $2,445. The lowest non-Essentia Hosptial was Fairview Lakes using Medica's PMAP at the price of $8,793.

Example Two: Identify the top 100 hospitals for minimal allowed amount for a given provider "Health Partners".

```
select top 100 hos.hname,
       pro.pname,
       min(cost.allowed) as minCharge
from dbo.Costs cost
inner join Hospital hos on hos.hid = cost.hid
inner join dbo.Product prod on prod.pid = cost.pid
inner join dbo.Offers as ers on prod.pid = ers.pid
inner join dbo.MCO as mco on ers.mid = mco.mid

where mco.mname = 'HealthPartners'
  and cost.drg_cd = '177'
group by hos.hname,pro.pname
order by minCharge
```

This query returned a similar result as example one but specifically for Health Partners. The lowest allowed amount was still for Essentia Fosston. However, the lowest for the Fairview health care system was for Fairview lakes product PMAP provided by Health Partners at $9,189 so about a $400 increase over Medica from the last example.

	hname	mname	pname	z_score_drg_code	count_drg
1	Essentia Moose Lake	Medica	MEDICAID IP	-0.81808899478173	759
2	Fairview Lakes	Medica	Medica PMAP	-0.811105715158145	746
3	Fairview Northland	Medica	Medica PMAP	-0.811105715158145	746
4	Essentia Brainerd	Medica	SENIOR CARE PLUS	-0.733541502259757	759
5	Fairview Lakes	HealthPartners	HP PMAP	-0.686556832796026	748
6	Fairview Ridges	HealthPartners	HP PMAP	-0.641048166155205	748
7	Essentia Duluth St. Mary	Medica	SENIOR CARE PLUS	-0.369989627006019	752
8	Essentia Detroit Lakes	Medica	SENIOR CARE PLUS	-0.246699386600242	759
9	Fairview Northland	HealthPartners	HP PMAP	-0.209748343032776	748
10	Essentia Virginia	Medica	SENIOR CARE PLUS	-0.0188233261396976	759

Fig. 3. Average z-scores across hospital, MCO, and product.

Example Three: Identify the top 100 providers for minimal allowed amount for Fairview.

```
select top 100 hos.hname,
       mco.mname,
       min(cost.allowed) as minAllowed
from dbo.Costs as cost
inner join dbo.Hospital hos on hos.hid = cost.hid
inner join dbo.Product prod on prod.pid = cost.pid
inner join dbo.Offers as ers on prod.pid = ers.pid
inner join dbo.MCO as mco on ers.mid = mco.mid
where drg_cd = 177
  and hos.hname like '%Fairview%'
group by hos.hname , mco.mname
order by minAllowed
```

This final example provided the lowest allowed amount for only hospitals in the Fairview health care system. Interestingly, Medica is the lowest for both Lakes ($8,793) and Northland ($8,793) locations but not for Ridges ($13,777) where the lowest allowed amount for Ridges is Health Partners at $9,508.

The above three examples are illustrated further in the Rshiny dashboard, discussed in section 6.3, along with many other useful subsets of information pulled from this database.

6.2 Result Two

Another result from this database includes comparing Hospitals and Providers by calculating the Z-Score of each DRG code and then taking the average across all Z-Scores (a function that returned a demean-ed normalized result).

Example One: Comparing the average allowed amount Z-scores per hospital.
Figure 3 shows the results from a query that calculates the z-scores for each DRG code and averages them across hid, mid, and pid. As you can see in Figure 3 Medica and Medicaid (PMAP is similar to Medicaid) will on average have the lowest allowed amount for any given Z-score. This result makes since due to how Medica and Medicaid were created to help the higher risk population (senior care) have an option for affordable health care. However, another interesting result from this query is that Mayo seems to have very bad results when compared to other hospital systems. Just looking at the table, Mayo Rochester on average will be 5 standard deviation away from the mean. If the allowed amount for a given DRG code has a normal (Gaussian) distribution, then this result should be near impossible. Looking deeper into this, we found that Mayo has the same prices for all providers regardless of product. However these findings were not significantly sound due to the fact that Mayo hospitals on average did not have a high number of DRG codes available. These results suggested that Mayo might use other methods to display information and we'll discuss this in the Limitations and Expansion sections.

Example Two: Comparing the average allowed amount Z-scores within the Essentia hospital system.
A large takeaway from the average Z-score results were that even though a given hospital was within a specific hospital network, that doesn't mean it will give you the lowest allowed amount on average. One example of this was in the Essentia Health System, using the commercial United Health Partners Product. In Figure 4, you can see that in Brainard, the average Z-score for a given DRG code is about 0.5, within one standard deviation. However, when you go to Duluth St. Mary's Hospital, you could be paying on average 1.4 standard deviations higher than the mean for a given DRG code on average.

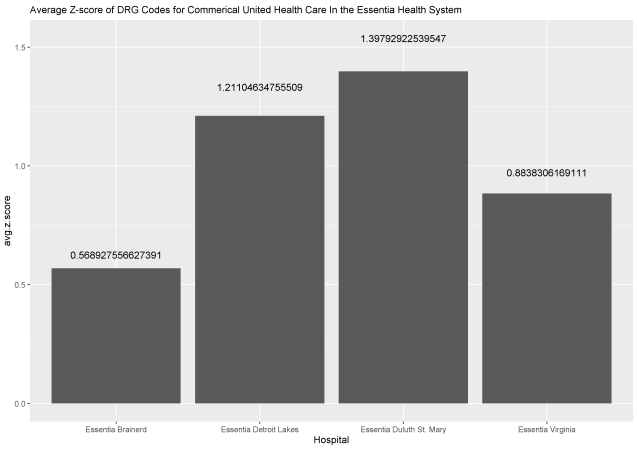

Fig. 4. Average Z-scores of DRG codes for commercial united health care in the Essentia hospital system.

6.3 Result Three

The third result this database assisted in, was implementing a mock-up Rshiny Dashboard. The statistical programming language, R, has a very useful dashboard implementation tool called Rshiny. The idea behind Rshiny is to create a dynamic interface that allows for a user to change the tables or graphs on the fly by entering specific information. This tool is useful in our case for many reasons. First, there are hundreds of DRG codes in our database and this allows a user to find specific cost information for any DRG code they enter. Also, this allows users to search the database by hospital, provider, or product, which can provide useful insights for that given individual. Lastly, this tool is useful for pulling in a large amount of hospital pricing data and yet creating generalized understanding for any particular person.

Figure 5 has an example of the Rshiny tool in action for someone who has Health Partners but doesn't know which product to buy. By typing in a DRG code that they either have or anticipate having and selecting the provider 'Health Partners' from the drop down and their hospital of choice Fairview Lakes, they are able to quickly see the table on the right update to the minimal allowed amount for their specific DRG code, provider, and Hospital.

7 LIMITATIONS

7.1 Database and data limitations

The extensive cleaning process needed for each raw data set reveled many idiosyncrasies, which although overcome for the purpose

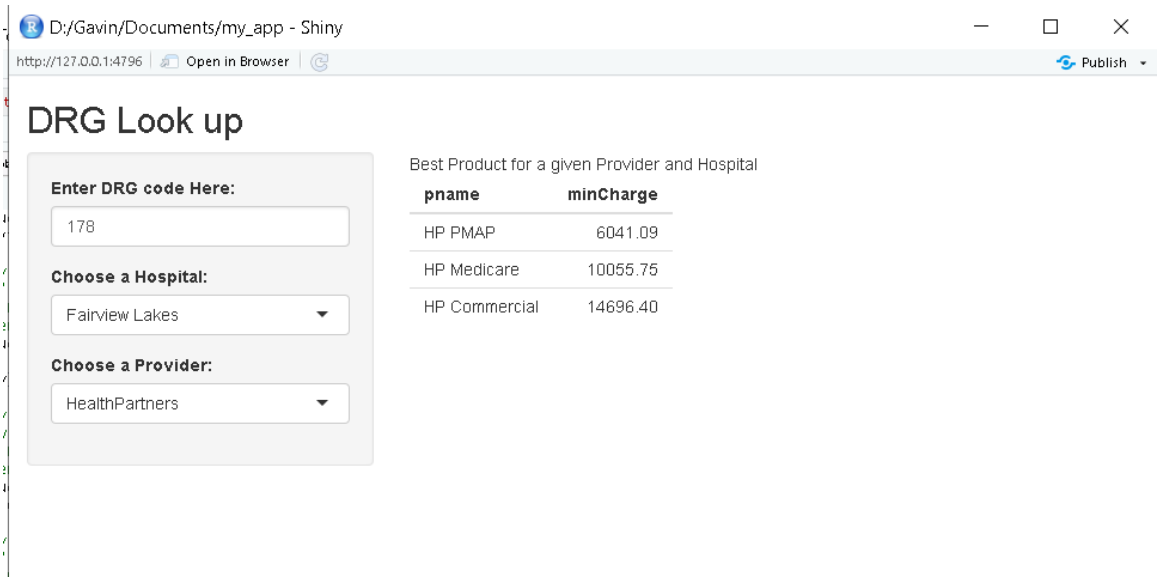

Fig. 5. Sample of Rshiny dashboard implementation utilizing this database.

of warehousing did result in some data loss and questions of data efficacy.

Of the hospital systems examined, several had unique data types that could not be included in our database. Per Diem rates were included only by Essentia, and average rates and uninsured data rates were included in some hospitals but not others. These rates were remove to preserve clean and usable data typing across the database. The database may be expanded to include more types of pricing data such as averages or Per Diem.

Upon examining the data from Mayo, we found that charges were the same for all MCOs and Products within a single DRG code. Put another way, any one given service at a given Mayo hospital only has one flat charge that is used for all providers. Whether this is actually how charges are calculated at these hospitals or if chargemaster data was retrofitted to an MCO spreadsheet is unclear.

The Essentia hospital system occasionally included more than one price for the same MCO and product combination. The purpose of these various costs was unclear, and disallowed by the structure of our database. To fit their information into our database structure, all but the first cost were removed prior to database integration. The removal of these values poses a limitation of the database structure, and the presence of these values poses a question of hospital compliance to the publishing rule.

Throughout the cleaning process we were surprised by how highly varied and often 'dirty' this hospital pricing data was. This lead to questions of hospital compliance to publishing guidelines - although all hospitals examined could be said to technically be in compliance, the presence of questionable and unintuitive pricing calls accuracy into question. Without insider knowledge or context from within the hospital systems it is impossible to say if this data was fully accurate.

7.2 Pricing Data for Consumer Use

Despite significantly more data available to potential patients, accurate data on what a patient is expected to pay is still out of grasp and utility is limited from a consumer standpoint. We now can say approximately what a hospital may charge an insurance provider for a particular DRG Code (assuming that rates have not change since they were last posted). However, the final bill to the patent will not come from the hospital; it is their insurance provider that will send the final charge. Each insurance plan has specific and often complex billing methods which are individualized to the insurance group, family, or person. The actual bill a patent is charged may depend on their co-pay, if they have a deductible, and if that deductible has been met so far in the billing year. Even individuals in the same insurance plan may have different deductibles based on a number of dependents.

An additional level of opacity for consumers is the nature of *how* this data is reported. A medical coding based approach is friendly to medical industry veterans but unknown to most consumers. Patents may be aware of what general procedure they are getting or had done, such as "head injury treatment", but they would be unable to assess if that would be coded as "Concussion", codes 088-090 depending on complication level, or indeed which complication level they have. Unfortunately alternative reporting methods can be more general and easily searchable only at the cost of accuracy, such as aggregating related codes and averaging them. Similarly, self diagnosis would be highly accurate but very imprecise.

The final and often unspoken limitation of pricing data as a tool for comparison and driving competition is the innate incompatibility of patient focused healthcare and a market based approach to purchasing care. Only rarely are individuals able to 'shop' for care in the first place, given the often emergency nature of many procedures. Meaningful choices can only be made before the care

is needed, and one can only make educated guesses at the type of care needed in the future. Additionally, cheaper care may not be better, and pricing data is only one aspect in a complex analysis of health options - distance, open beds, availability of specific and perhaps specialty procedures all must play a roll. DRG Codes by their very nature do not include preventative care, which has been shown to be the most cost effective way to ensure longevity and health [Maciosek MV 2010].

8 EXTENSIONS AND NEXT STEPS

8.1 Additional Data

The main extensions to this project would be to add more hospitals to the database. This would allow for more in depth analysis of Z-scores and allow for analysis by state if we obtain hospitals from other states. Other information that would be beneficial to include in the database would be new codes (such as CPT codes) and creating a crosswalk between CPT and DRG codes. This would be useful because some hospitals like Mayo have a lot of codes in CPT which we could include because we didn't have a good way of combining CPT codes with DRG codes. However, a crosswalk would allow for the inclusion of these codes and others to analysis more procedures. Finally, we could expand our application of the data to a public website and include other useful insights such as average allowed price over a given set of DRG codes, average change in price over time, or given a more general diagnosis or procedure, what would be a likely price at a given hospital.

8.2 Expansion of Datamart Model

The model could also be extended to include both geographic information and market information as described below.

Geographic Information (Hospital Attribute) - The geographic location of a hospital can influence reimbursement rates. Generally, rural hospitals are able to charge a higher rate than urban hospitals. There are many reasons for this, including–but not limited to–hospitals having a monopoly on the service market. Similarly, hospitals require higher reimbursement to pay doctors a higher salary to live in rural areas.

Market Information (Product Attribute) - The price of a service for a hospital depends on the product of the insured individual. Figure 5 shows Fairview Lakes' prices for DRG code 178 for the three HealthPartner products. Pre-paid Medical Assistance Program (PMAP) is a Medicaid program for low-income Minnesotans administered by the Minnesota Department of Human Services (DHS). Medicare is a program that is administered by CMS for seniors (aged 65+) and for individuals with disabilities. Finally, commercial rates are for individuals covered by their employer or who purchase coverage through the Individual market (e.g. purchased through Minnesota's state-based exchange called MNSure). Hospitals typically give discounts to MCOs for Medicaid products. Medicare rates are set by the Federal Government through the IPPS and hospitals will usually negotiate reimbursement from MCOs as a percentage of IPPS. Finally, commercial products generally pay between 151% - 222% to the Medicare IPPS [KFF 2020]. The current model does not account for this given the time involved and market-knowledge required to identify the correct market of each product.

9 CONCLUSION

This project illustrated that, although feasible, warehousing such diverse and distributed data can only be done with extensive management costs. Any automation, such as for yearly cost updates, would be limited in scope and require routine checks by knowledgeable personnel. Bespoke cleaning and assessment would be needed for every new hospital system added to the database. Standardization of content and format through clearer and more explicit publishing guidelines could streamline this process, as could creation of a centralized directory of raw data or links to raw data.

This begs the question, would the investment in creating a larger, perhaps nation wide, database of negotiated healthcare costs be worth the information gained? Although this data may have limited benefit to consumers, there is clear benefit to other actors in the health care industry. Both hospitals and insurance providers can leverage this information in cost negotiations, and legislatures can use it to craft more accurate and applicable policy. Whether this data will lead to lower prices as hospital pricing becomes more transparent has yet to be seen.

10 ACKNOWLEDGEMENTS

The data for this database was sourced from publicly available websites. However, each hospital system includes this information on different locations within their websites. The three Minnesota based hospital systems used in this project were Mayo, Essentia and Fairview. The pricing data for Mayo was obtained on location specific websites [Mayo 2021]. Whereas for Essentia and Fairview, there was a central location which holds links to each of the locations data. [Essentia 2021; Fairview 2021].

11 DIVISION OF WORK

The division of work for this project was dispersed evenly throughout the four team members of group 11. All members contributed to each step in the project process. As noted, a large portion of the project entailed data mining, cleaning and standardizing. Therefore, the three hospital systems were each assigned to a member. Abby worked on Mayo, Destiny worked on Essentia and Gavin worked on Fairview. During this time, Tyson worked on obtaining and cleaning the necessary data from various websites for the DRG codes and descriptions. Another large part of the project was working with Azure server management, which Destiny set up and managed for the group. Overall, many areas of the project were a group effort.

REFERENCES

CMS. 2021. PQDC. https://data.cms.gov/provider-data/dataset/xubh-q36u

Essentia. 2021. Price Estimates and Standard Charges. https://www.essentiahealth.org/patients-visitors/billing/price-estimates/

M Health Fairview. 2021. Billing Information and Financial Services. https://mhealthfairview.org/billing/patient-billing-financial-services

HollandKnightHealthcare. 2021. *Price Transparency in Hospitals – Is Hospital Pricing Data a Protected Trade Secret?* https://www.hklaw.com/ja/insights/publications/2021/09/price-transparency-in-hospitals-is-hospital-pricing-data

ICD. 2022. 2022 Diagnostic Related Groups List (MS-DRG v39.0). https://www.icd10data.com/ICD10CM/DRG

KFF. 2020. How Much More Than Medicare Do Private Insurers Pay? A Review of the Literature. https://www.kff.org/medicare/issue-brief/how-much-more-than-medicare-do-private-insurers-pay-a-review-of-the-literature/

Flottemesch TJ Edwards NM Solberg LI. Maciosek MV, Coffield AB. 2010;. Greater use of preventive services in U.S. health care could save lives at little or no cost. *Health Aff (Millwood)* 29(9) (2010;), 1656–60. https://doi.org/10.1377/hlthaff.2008.0701.

Mayo. 2021. Locations. https://www.mayoclinichealthsystem.org/locations/

Northfield 2021. *Price estimate / price transparency: Northfield Hospital and Clinics*. Retrieved Dec 15, 2021 from https://www.northfieldhospital.org/price-estimate-price-transparency