Final Research Paper

Nicholas Wolpers

Colorado State University Global

MIS581 – Capstone: Business Intelligence and Data Analytics

Dr. Justin Bateh

Due June 4, 2023

Table of Contents

Abstract

Hospitals and healthcare services in the United States are facing a large and escalating problem of runaway drug prices and heightened drug spending. Incorrectly-applied pricing is a large but, according to literature review, currently uncounted aspect of drug spend that has a detrimental impact in the context of rising drug prices. The present work offers an introduction to the dynamics of the pharmaceutical marketplace for US hospitals, describes an author-designed SQL-based analytical technique using inductive statistics to identify pricing disparities, and concludes with elements of a case study for one hospital system (CommonSpirit Health) which has used this method to increase their recovery of overspending dollars to a conservative $3.3M over the past two quarters to date, representing an almost seven-fold increase in savings over prior quarters.

Introduction

Hospitals and consumers of healthcare services in the United States are facing a large and escalating problem of runaway drug prices. As one report by the American Hospital Association and other entities puts it, “Continued rising drug prices, as well as shortages for many critical medications, are disrupting patient care and straining hospitals’ budgets and operations” (*Recent Trends*, 2019). The issue of rising costs in the inpatient hospital setting can be attributed to some degree to the complexity of the drug purchasing marketplace. This complexity can be expressed in several ways, with one such mechanism being price mistakes, whereby the incorrect application of higher prices by vendors causes overspending and the overextending of hospital drug budgets and other adverse effects. Such errors are curable to a certain extent, provided the error is identified in requisite time. This work will examine an analytical methodology to capture overspending on pharmaceutical purchases using inductive statistical techniques via a custom SQL-based solution.

To identify the motivation and application of such a technique requires a fair amount of background on the drug purchasing marketplace for the hospital setting. In the modern hospital, the bulk of purchasing is typically done through a group purchasing organization (GPO). A GPO is an entity that “helps healthcare providers — such as hospitals, nursing homes and home health agencies — realize savings and efficiencies by aggregating purchasing volume and using that leverage to negotiate discounts with manufacturers, distributors and other vendors” (HSCA, n.d.). These discounts come in the form of dozens to hundreds of GPO contracts made available to member organizations, which may in turn cover thousands to tens of thousands of products.

Purchase transactions are realized via accounts established with a pharmaceutical distributor (also known as a wholesaler), through accounts commonly termed “GPO accounts”. Voluntary involvement with government programs may result in additional account types with different dynamics. One example is the “340B” account type, tied to the 340B Drug Pricing Program which “enables safety net health care providers to purchase outpatient drugs at discounted prices” (*Drug Spending*, 2023). According to the governing entity for the program, the Health Resources and Services Administration (HRSA), “enrolled hospitals and other covered entities can achieve average savings of 25% to 50% in pharmaceutical purchases” (American Hospital Association, 2023). Another is the “wholesaler acquisition cost” (WAC) account type, which is based around the manufacturer’s list price. This can be likened to the manufacturer's suggested retail price (MSRP price) for purchasable drugs, and is often the highest-priced channel for drug acquisition despite its status as the benchmark for pricing throughout the pharmaceutical market.

Each account type comes with their own dynamics and restrictions. To complicate matters further, even within these account types there are classifications for different pricing structures depending on the setting or channel for which the purchased drugs are to be used, referred to as “Class of Trade”. Examples of Class of Trade span settings as varied as retail pharmacies, inpatient use, outpatient/clinic use, and so forth, each with both state and federal legislation regulating their definition and appropriate classification. With 100+ possible designations for Class of Trade, there are severe challenges in managing a consistent methodology for assigning and maintaining the Class of Trade to accounts (Prescription Analytics, 2022). To complicate matters further, there are also subdivisions within the account types, with the 340B account type having nearly a dozen sub-types dependent on the purchaser’s

federal classification within the 340B program; each come with varying restrictions or privileges. And while the specific reasons are beyond the scope of this paper, most hospitals that qualify for reduced drug pricing via the 340B program are mandated to spend a portion of their drug budgets buying pharmaceuticals on WAC accounts at the non-discounted price.

Compounding the classification issues listed above, there are multiple layers of pricing dynamics relating to the *contracting* of drug purchases. A typical benefit of a GPO is the large swath of pre-negotiated drug contracts across many account types and classes of trade, numbering from dozens of contracts to hundreds, for purposes of securing prices lower than the default (and typically most expensive) wholesale acquisition price. Savvy purchasers may opt to pursue individual contracts with drug manufacturers for even better deals than offered by the GPOs, in exchange for terms that might stipulate certain purchase volumes over time (bulk purchasing) or greater marketshare (exclusivity arrangements for certain baskets of drugs). There may even be multiple tiers to a given contract, with multiple levels of pricing concessions from manufacturers dependent on set targets. GPO contracts and individual contracts, like any contract, also have set durations, and there may even be fallback contracts which are meant to activate once primary contracts expire. Against this backdrop of a dynamic marketplace with fluid, multilayered pricing and legal restrictions, one must be made aware the size of this market is an estimated $603B per year (Parasrampuria & Murphy, 2022).

The pharmaceutical distribution ecosystem in the US is considered a consolidated market – the "big three" wholesalers (AmerisourceBergen, Cardinal Health, and McKesson) maintain a collective 92% marketshare (Hisey et al, 2019). In addition to distributing, these companies also "purchase and take legal ownership of pharmaceuticals and manage both inventory and credit

risk, [allowing] manufacturers to focus on their core competencies of pharmaceutical drug development and manufacturing" (Hisey et al, 2019).

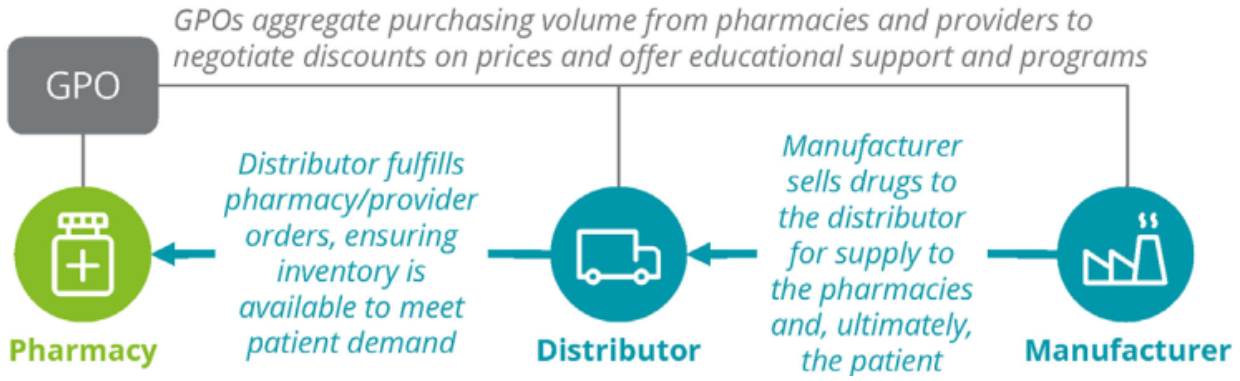
Figure 1

Pharmaceutical wholesalers play a large role in the drug marketplace

Note. Adapted from *The role of distributors in the US health care industry: 2019 report* by Hisey et al, 2019, p. 4. Copyright 2019 by Deloitte Development LLC.

Figure 2

Simplified diagram of drug supply chain



Note. Adapted from *The role of distributors in the US health care industry: 2019 report* by Hisey et al, 2019, p. 4. Copyright 2019 by Deloitte Development LLC.

There is no doubt that wholesalers serve multiple value-adding roles in the pharmaceutical supply chain, while also playing an intricate role in the handling of transactions. As a Deloitte reports succinctly puts it, “A leading example of how distributors support ecosystem efficiency beyond moving product is by managing contracts and facilitating chargebacks. Distributors manage contracts between GPOs and manufacturers, enabling GPOs to use the bargaining power of their members to earn discounts. Distributors monitor sales between these ecosystem players and facilitate the payment of chargebacks from the manufacturer in the amount of the difference between the prices negotiated by the GPOs and the wholesaler acquisition cost (WAC)” (Hisey et al, 2019).

Objectives

The preview of the complexity behind the determination of contract management and pricing provided above prompts the question: what about when things go wrong? Incorrectly applied pricing is a large but, according to literature review, currently uncounted aspect of drug spending that has a presumed detrimental impact in the context of rising drug prices.

The very concept of “incorrectly applied pricing” has been identified by several terms across multiple parties in the pharmacy landscape: price arbitrage, price misalignment, price disparity, price verification, and price integrity, to name a few. Compounding the presumed negative impact are the restrictions around redress. The application of pricing corrections is typically asymmetrical: if drug manufacturers find pricing mistakes to their detriment, they reserve the right to apply corrective billing for a span of time up to 12 months prior to discovery.

For hospitals which have found price mistakes to their detriment, there is a maximum span of three months to apply for retroactive corrective billing. To correct for this discrepancy, hospitals require advanced analytic models to discover price mistakes quickly and accurately so as to correct excess spending in due time.

As noted within other industries like hoteliers, "All price disparities are damaging and you must identify them, correct them and create a structure so that they don't appear again in the future" (Delgado, 2017). In that vein, this report will discuss one such analytical intermediation method designed and applied for CommonSpirit Health, a not-for-profit 100+ hospital system spanning the continental US, and the $2.6M in overspending it has recouped across a six-month timeframe. After all, pricing affects resource allocation, so by addressing recoverable overspending, hospitals can act to mitigate elements contributing to drug budget inflation and restore the focus on the patient and their mission.

Overview of Study

This report offers an introduction to the dynamics of the pharmaceutical marketplace for US hospitals, describes an author-designed SQL-based analytical technique using inductive statistics to identify pricing disparities, and concludes with elements of a case study for one hospital system (CommonSpirit Health) which has used this method to increase their recovery of overspending dollars to a conservative $3.3M over the past two quarters to date, representing an almost seven-fold increase over prior quarters.

Research Questions and Hypotheses

In devising analytical means to address the issue, having a solid hypothesis helps to frame the matter and communicate the distinct objectives of the research. Generating the hypothesis in question should ideally reflect part of the premise: whether an analytical technique or process can convey sufficient information to identify incorrectly-priced pharmaceuticals *without* having a source of truth regarding the pricing, while the market possesses many different layers of complications. Paid consultant research (McKinsey) had indicated it may amount to nearly one million per year for my employer, CommonSpirit Health, yet the specific methodology used in determining this figure was not shared.

As the rate of occurrence of such pricing mistakes is unknown at the outset, it would be difficult to craft hypotheses which target a certain rate of pricing mistakes to be captured. A more appropriate measure for such a hypothesis, if one is to use dollar-denominated value, is to meet or beat prior achieved results. Formerly, the consulting arm of a large national group purchasing organization was retained to monitor and correct pharmaceutical pricing inaccuracies with quite generous terms, including a percentage of the amount recovered. Even with this advantage, the recovered dollars were estimated to be on the low side (see **Table 1**). If that is to be the baseline which an analytical technique or process must exceed, then the hypothesis can possibly be defined as such:

> H_0: There is no difference in the results of the new analytical process on the average quarterly recovery (in dollars) from the correction of pharmaceutical price mistakes by the extant process using an outside consulting company.

H_a: There is a difference in the results of the new analytical process on the average quarterly recovery (in dollars) from the correction of pharmaceutical price mistakes by the extant process using an outside consulting company.

Table 1

The amount of "price verification" funds recovered in prior periods when done by outside consultant group

Fiscal Period	Value Recovered
FY22 Q2	$23,605
FY22 Q3	$377,526
FY22 Q4	$76,168

The statistical basis for rejecting the null hypothesis and accepting the alternative hypothesis can be derived by computing the analysis of variance (ANOVA) and assessing the values of the p-value of the results of the F-test. As defined by Elliott and Woodward (2016), ANOVA is used to compare means across groups or to compare three or more repeated measures (dependent samples). A one-way ANOVA is an extension of the independent group t-test where there are more than two groups, with essential assumptions being that data is normally distributed with equal variances across groups, and independent samples (Elliott & Woodward, 2016). Yet thankfully, "the ANOVA is robust against moderate departures from the assumptions of normality and equal variance (especially for larger sample sizes)" (Elliott & Woodward, 2016). The null hypothesis for such an analysis is that the means of all the groups are equal, with the alternative hypothesis framed such that at least two means are not equal. A low p-value for the F-test is evidence for rejecting the null hypothesis (Elliott & Woodward, 2016).

It should be noted that this level of aggregation (dollars per quarter) presents a challenge with regards to the sample size, but not an insurmountable one. The limitation is due to the reporting standard of the consulting group, which only communicates a single value (total dollars recovered minus fees) once per quarter. A side benefit of the new analytical process will be greater transparency into the issue at a greater granularity.

Literature Review

Part of what makes the problem this endeavor addresses interesting in the first place is the complexities of the market means there is no known source of truth for what the price *ought* to be from any information source, making every single exception on the report an exercise of inductive reasoning. Purchasers of identical pharmaceutical products with certain characteristic have one characteristic distribution, and other purchasers with other characteristics have another (perhaps different) distribution. How to find actionable outliers in such a scenario?

The statistical tool which was most inspirational to the chosen approach comes from Bayes inductive logic and the Naïve Bayes classifier methodology, in which probabilistic reasoning is expedited by presuming independence among various predictors. Even a brief review reveals there is a whole family of algorithms that fall under this umbrella, which leads to many published case studies and papers to select from when citing the utility of the approach. Karakostas (2016) helpfully frames his usage of Naïve Bayesian models as "allow[ing] us to infer dependent probabilities of events, from other events that are either easier to detect or to

predict", as is the case here, and also demonstrates a prototype using database architecture to predict outbound flight delay events based on inbound flight delays.

Several difficulties present themselves in the course of this literature review, especially concerning the depth and breadth of economics research. While many published papers in the economics space *do* concern price discrepancies, often these approach the matter in the context of open markets and taking advantages of such price discrepancies (as in the stock markets). No papers were located concerning more complex or semi-open markets, where there should be an established price but it is unknown, and also there are regulations regarding which actors can access certain markets (e.g. only some entities are granted access to preferential pricing) and also steep penalties for getting it wrong. A hospital that buys more products through the discounted channels (like 340B) than it has demonstrably earned is subject to paybacks and steep fines, government and pharmaceutical manufacturer audits and mandatory improvement plans, and even complete ejection from all discount plans, among other legal ramifications. In a similar but arguably less punitive and more opaque process, pharmaceutical manufacturers which are known to have misconstrued the financial situation of their products in non-public government filings are penalized by steeper discounts on said drugs through the discounted purchasing channels for one or more quarters. It is not unheard of that drugs which have wholesale costs in the thousands are sold for pennies due to such dynamics (called "penny buys" in industry jargon). Sadly, no journal-published articles were found which discuss such dynamics in detail.

The most useful series of papers were from certain authors like Kathleen Iacocca whom have expressed an interest in pharmaceutical pricing over time across multiple contexts. One which will be leaned on heavily for this report concerned a machine learning model for

predicting the prices of drugs using a number of predictors. Note that machine learning is not even the methodology of the technique under discussion but more of a competing methodology, yet the paper shows its value by its uncovering of the complexities of the space and even its failures (in both data sourcing due to its heavily proprietary nature and in the predictive strength of its models). The highest-scoring of the machine learning methodologies Iacocca had explored to predict drug prices returned an R-square value of 0.54 (Iacocca & Vallen, 2021).

Also of interest was a paper by Bahl et al (2007) concerning the creation process for a pharmacy dashboard which discussed the practical considerations and notably the fluctuations and dynamics of drug pricing on and within such visualizations. This paper touches on the subject from the direction of cost control, which is a shared objective because the technique under discussion is intended to avert and allow correction for unwarranted expenses on drugs above negotiated rates. It also had relevant discussion of how the data warehousing process was approached, which will be leaned on for similar discussion in the report to come.

Non-profit research and work in this space is also immensely valuable, as ideally it stands apart from industry-sponsored or hospital-association-sponsored research. One report by the RAND corporation concerning US drug pricing research discussing price comparisons and price indicies proved very intriguing, especially with one focus area being different methodological steps and assumptions with different levels and volume weights calculated in different ways (Mulcahy et al, 2021). The RAND report is highly detailed and does well to approach their stated objective of “providing health care decisionmakers, practitioners, and consumers with actionable, rigorous, objective evidence to support their most complex decisions” (Mulcahy et al, 2021).

All said, it seems like an endless number of references avail themselves with sufficient time and research, yet with none addressing the precise problem at hand. This is both a promising but also a difficult situation, as each new resource offers a new sliver of insight on the present issue but only when stacked with multitudes of others in just the right configuration.

Methodology

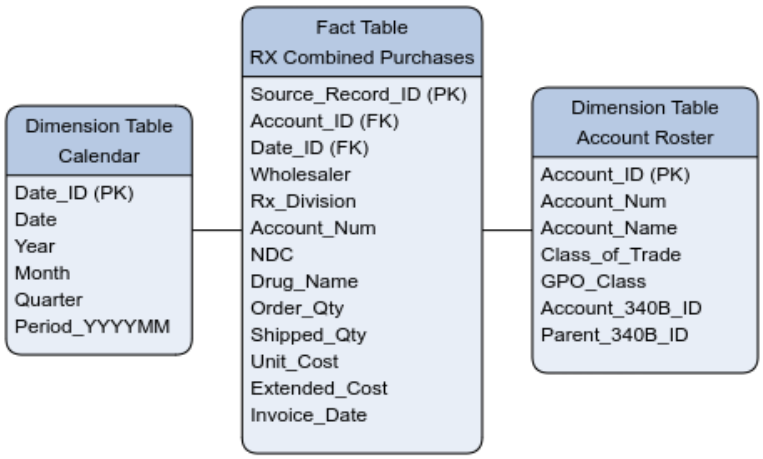
The dataset for which this analytic technique is demonstrated comes from a proprietary drug procurement relational database maintained by the System Pharmacy staff of CommonSpirit Health. The dimensional model for this relational database management system is in the form of a star schema (see **Figure 3**). The primary dimension tables in use are an account roster dimension and a date dimension, with the fact table comprised of line-item invoices from drug purchases. This fact table spans multiple vendors – in truth it spans as many vendors as can supply sufficiently detailed data at the grain of a "line-item invoice" on a routine basis. Currently, this relational database covers five vendors (McKesson Drug, Amerisource Bergen, McKesson Medical-Surgical, FFF, and Civica Rx) with a row count in the hundreds of thousands of invoice line items per month, and an average expense of roughly 81 million dollars per month.

The fact table contains line-item detail for purchases, returns, and credit/rebills (post-facto pricing corrections, or records of the transfer of purchases from one account to another for reasons beyond the scope of this paper). Certain drugs are exclusive to one or another vendor, but there is a large overlap in the catalogs of most vendors when it comes to purchasable drugs,

as well as differing price points. Crucially, each line item invoice presents the Vendor, the account number, the National Drug Code (NDC) of the purchased drug, the unit price, and the net extended price for each purchase.

Figure 3

Simplified diagram of the Combined Drug Invoice star schema



An NDC number is a Food & Drug Administration (FDA)-assigned identifier of all drugs manufactured, prepared, propagated, compounded or processed for sale in the U.S. by all drug labelers (National Drug Code Directory, 2022). These drugs are identified and reported using a unique, three-segment number, commonly represented as either a full NDC (11-digits), or a partial NDC (10-digits or fewer). Furthermore, "the NDC Directory contains product listing data submitted for all finished drugs including prescription and over-the-counter drugs, approved and unapproved drugs and repackaged and relabeled drugs", demarcating its intended use as a

universal product identifier for human drugs (National Drug Code Directory, 2022). As of April 29, 2023, the FDA-provided downloadable NDC database file spans 108,000 line items, updated daily. For analytical purposes, an NDC can be considered analogous to a "stock keeping unit" or SKU number, those unique number combinations used by retailers to identify and track products.

The three segments of the NDC communicate certain information about each product. The first set of numbers (5-digits for a full NDC) identifies the labeler of the drug. Note that a labeler "may be a manufacturer, including a repackager or relabeler, or the entity named on the product label" (National Drug Code Directory, 2022). The second set of numbers (4-digits for a full NDC) represents the product code, which identifies the specific strength, dosage form (i.e., capsule, tablet, liquid) and formulation of a drug for a specific manufacturer" (RS & F Healthcare Advisors, 2019). The last set of numbers (2-digits for a full NDC) represents the package code, which identifies package sizes and types and is set by the drug's labeler. Accuracy regarding the proper NDC is crucial for the analytic technique to function.

Pricing, as any economist of the last century will tell you, communicates a lot. While many prices in the pharmaceutical marketplace are negotiated, many more must be bought on the open market at an ever-evolving list price. And drug list prices, "which form the basis for prices paid throughout the pharmaceutical supply chain", are set by the manufacturer (Iacocca & Vallen, 2021). In the report titled *Using Analytics to Gain Insights on U.S. Prescription Drug Prices*, authors Iacocca and Vallen (2021) make it clear that "given the disjointed nature of data on pharmaceutical drugs and drug pricing, there are few inquiries in the literature that explore the manner in which multiple variables directly and interactively impact pharmaceutical prices". Furthermore, they attest quite early in their article that "much of the research that explores the

factors underlying pharmaceutical list prices reveals findings that are inconsistent at face level (e.g., the entry of generic drugs can result in branded drugs increasing or decreasing in price)", and that "financial transactions are more complicated, with the prices different entities pay varying widely throughout the supply chain" which this writer can corroborate (Iacocca & Vallen, 2021).

The primary technical tool for accessing, processing and computing statistical values for this technique is T-SQL, here via Microsoft SQL Server 2019 and Microsoft's database management systems. Statistical measures of central tendency are captured using the standard functions provided with MS-SQL, operationalized through a sequence of custom scripts. Other tools include spreadsheet software (e.g. Microsoft Excel, Google Sheets) for disseminating and relaying actionable data to staff, in part due to their versatility and universality. The median and mean, and a user-specified threshold of distance from the median or mean that is expressed as a percentage. This threshold sets the boundary line for what is normal price variation for a facility with a given set of attributes. These attributes are the combination of all known attributes that determine the prices to be paid. For instance, a hospital purchasing *Drug A*, having a certain Class of Trade and 340B designation, should pay a comparable amount to a peer hospital that also buys *Drug A* and has an identical Class of Trade and 340B designation. Such data is extent in the data set at a latent level, and needs special algorithms to pull to the surface and make evident where price discrepancies are occurring and for what drugs.

To actively utilize the price integrity technique, the pharmacy information technology (IT) staff runs extracts from the pharmacy system on the first day of the month that pull all drug purchase transactional data from the previous month. This data is then processed and mapped to

the dimensional structure through a series of scripts. Once the data is converted, a series of quality checks are run to look for potential errors in the data. After successful completion of those checks, the data is uploaded into the warehouse. Said data is then extricated into a spreadsheet featuring pivot tables for review by system pharmacy purchasing team staff. The resulting data set is in the form of an exceptions report, which is quantitatively determined using inductive statistics. This report then goes to human reviewers, the ones who will be transmitting the information to vendors or wholesaler for correction, thus one could more properly describe the entire process as a combination of inductive and qualitative judgment.

Limitations

Like any tool, especially one which is developed using proprietary data, it is difficult to determine precisely how much of the results are generalizable. In addition, as the method of capturing exceptions relies on inductive statistics, there is a propensity for false positives. The root cause of these false positives appears to stem from two sources: errors in the rostering of accounts (the prices from misclassified accounts can skew the distribution for other account types), and faulty generalizations (prices which may be specific to only one facility or set of facilities are presumed universal when they are not). In general, the practice of a human review process resolves a large proportion of such errors, and the default behavior of the human reviewers has been to apply the most scrutiny to line items which indicate the highest return of investment. The rate of false negatives using this method is unknown, though it can confidently

be stated that 100% of all pricing corrections for the past two financial quarters have been from the use of this tool.

Ethical Considerations

In regards to ethical considerations for the type of data that has been chosen, special considerations have to be made for the proprietary nature of the dataset. While the data for this application is sourced from hospitals, it does not contain Protected Health Information (PHI). The HIPAA Privacy Rule is a standing law that provides protections to US denizens for personal healthcare-related data, but none of the data in question falls under that scope. Rather, this data is best defined as “proprietary” data, indicating any information that discloses business transactions, market positions or trade secrets. Such data is deemed sensitive and not disclosed to the public, so special handling is still warranted.

This data is granted for use for the purpose of this paper because of the author’s role as the developer of the tool, but the information in question is still owned by the organization (CommonSpirit Health) itself. Disclosure of such data could lead to damage to partner relations, harm future negotiations, and possibly fall afoul of various non-disclosure agreements between specific vendors, even resulting in legal liability. Regardless of the fact that the data is not reflective of individuals persons, it is still incumbent upon the researchers to take data security seriously after being granted use for the purpose of writing this report.

Findings

The core of the report had identified the purchasing outliers among the hospitals and their accounts which had purchased drug at wildly varying prices, the price distribution (high mark, low mark, average, median) of spend from similar hospitals or clinics and functional detail such as contract identifiers and invoice numbers. A recovery estimate termed the “impact benchmark” was produced using the difference between the recorded extended cost and a recalculated extended cost using the median price from the observed distribution. This value would allow users to prioritize the highest-value fixes from among the batch. A typical one-month data report might consist of 1,500 line items spanning 2,500 invoices, with the sum of the difference from the median unit price (indicating median recoverable dollars) totaling more than $500,000. Users voiced comments that there was always more than could even be followed up on in the course of a routine work week.

After emerging from the prototype stage the monthly report grew to have over 40 fields, with over half arising from user request. Later enhancements led to nearly a dozen fields pulling from various benchmarks to “paint a portrait” of the pricing landscape; fields for algorithmically prompting whom to approach for redress (wholesalers or manufacturers, etc.) derived from certain indicators; a summation of the extent of the pricing problem when extended across product lines (e.g. summing recovery estimates across multiple groups of NDCs); more extensive product information; and clearer manufacturer identification to indicate recurring problems. Initially the threshold for report inclusion was set to 25%, meaning that a given unit price had to exceed the inductively-established median price by 25% or over before inclusion in the report

would be warranted. Later this would be relaxed to 10% after months of positive outcomes. The savings values achieved by using this methodology can be seen in **Table 2** and **Figure 4**, starting from Fiscal Year 2023 Quarter 1 and onward.

Table 2

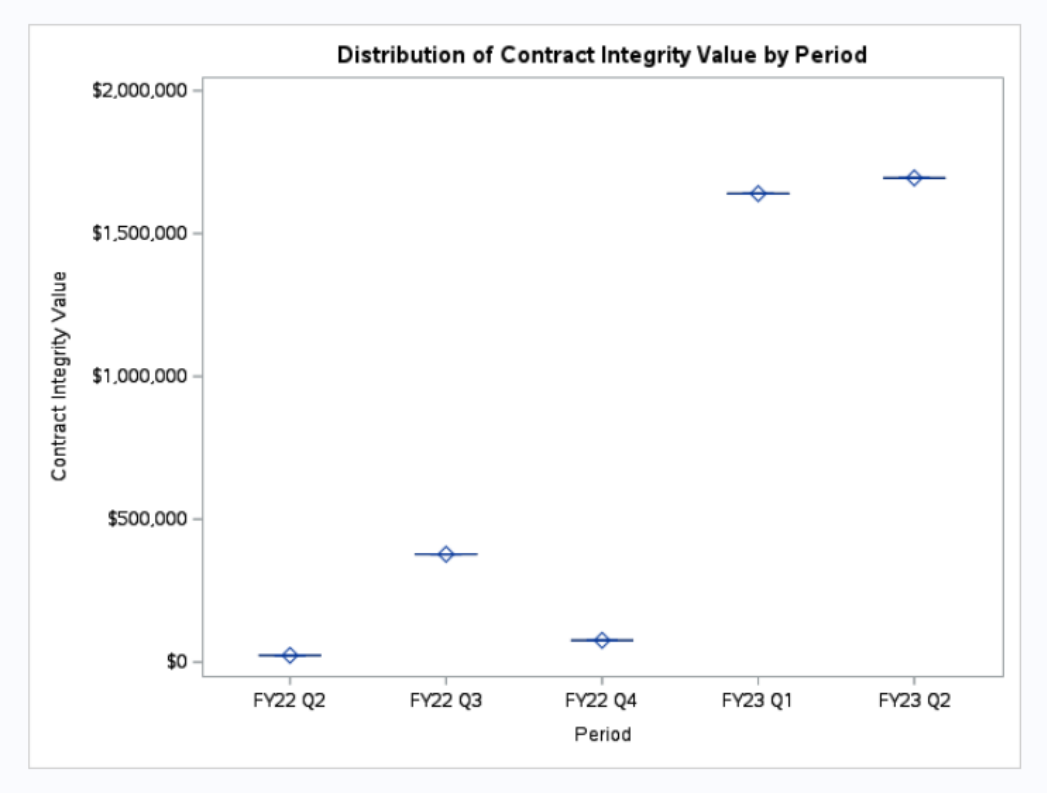
Total recovery dollars from overspending on pharmaceuticals over 15-month timespan

Fiscal Period	Value Recovered
FY22 Q2	$23,605
FY22 Q3	$377,526
FY22 Q4	$76,168
FY23 Q1*	$1,639,694
FY23 Q2*	$1,694,592

Note. Quarters utilizing new methodology marked by asterisks.

Figure 4

Contract Integrity savings over 15-month period, spanning both methods



Regarding the testing of hypotheses, documentation from the statistical software vendor suggests that while a one-way analysis of variance (ANOVA) *can* be used with unbalanced data (data with an unequal number of observations for every combination of classification factors), the dataset was in fact converted to a balanced one prior to operation (SAS Institute Inc., 2013). This was done by removing the earliest recorded datapoint for the consultant-led recovery ($23,605 for Fiscal Year 22 Quarter 2) which happened to be the lowest recorded value of the entire dataset, effectively balancing for a more favorable comparison between the two datasets on the whole. The results of this procedure can be viewed in **Figures 5-7** on the following pages.

Figure 5

One-Way ANOVA results and box plot for mean savings by recovery process using SAS Studio

Dependent Variable: Value

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2.0744526E12	2.0744526E12	88.43	0.0111
Error	2	46915217284	23457608642		
Corrected Total	3	2.1213678E12			

R-Square	Coeff Var	Root MSE	Value Mean
0.977884	16.17313	153158.8	946995.0

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Price Verification M	1	2.0744526E12	2.0744526E12	88.43	0.0111

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Price Verification M	1	2.0744526E12	2.0744526E12	88.43	0.0111

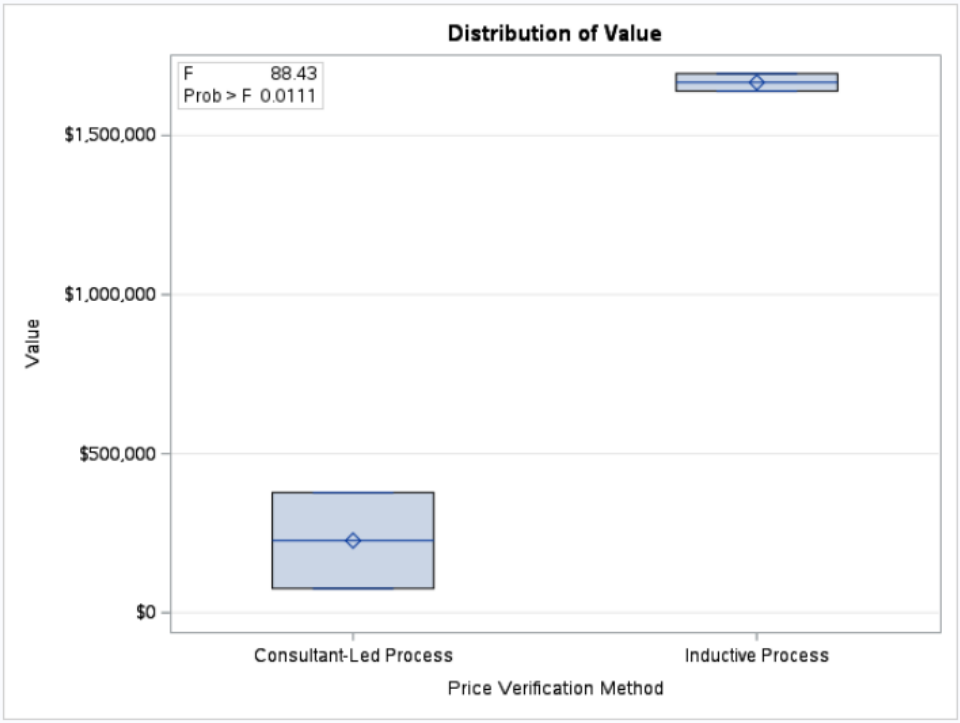


Figure 6

Welch's ANOVA for savings value for results between the two approaches

Levene's Test for Homogeneity of Value Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Price Verification M	0	0	.	.	.
Error	0	0	.		

Welch's ANOVA for Value			
Source	DF	F Value	Pr > F
Price Verification M	1.0000	88.43	0.0589
Error	1.0663		

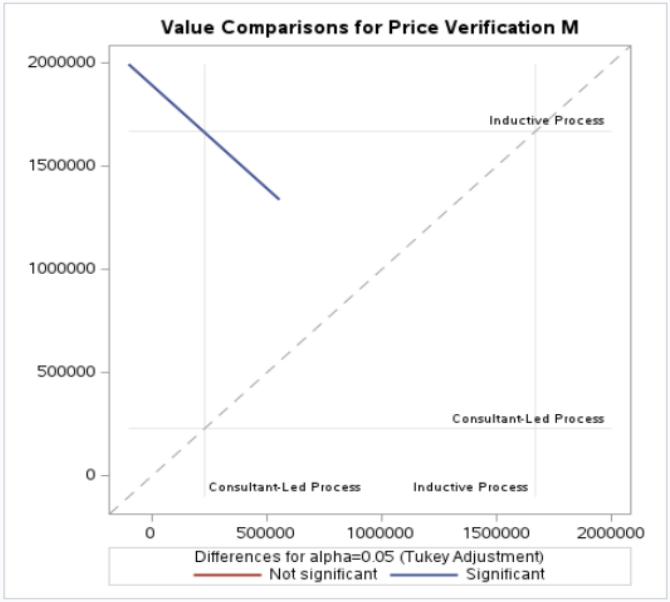
Level of Price Verification Method	N	Value	
		Mean	Std Dev
Consultant-Led Process	2	226847.00	213092.285
Inductive Process	2	1667143.00	38818.748

Figure 7

Least Squares means and value comparison between the Consultant-Led Process and the Inductive Process

Least Squares Means
Adjustment for Multiple Comparisons: Tukey

Price Verification Method	Value LSMEAN	H0:LSMean1=LSMean2 Pr > |t|
Consultant-Led Process	226847.00	0.0111
Inductive Process	1667143.00	



As confirmed by Elliott and Woodward (2016), a low p-value for the F-test is evidence for rejecting the null hypothesis in this situation. This finding supports the proposition that the inductive method is much improved over the consultant-led option. In an ideal scenario the two approaches could be measured head-to-head, yet as there were no facts or measures provided by the consultant-led process apart from dollar-denominated results, a dollar-denominated comparison of each method's results is as far as this comparison can go.

Conclusion

In summary, this report has discussed the acceleration of overspending and the complexities of the current US pharmaceutical market, a novel analytical methodology to capture overspending on pharmaceutical purchases using SQL-based inductive statistical techniques, and the application of this methodology by a leading healthcare system to identify and recapture excess spend. This hospital system (CommonSpirit Health) has achieved conservatively-valued savings of $3.3M over the past two quarters to date, representing an almost seven-fold increase over average prior quarters utilizing alternative means of correction. This is statistically supported using a One-Way ANOVA and F Test, with the low p value there (0.0111) indicating the null hypothesis may be rejected in favor of an alternative hypothesis that the new analytical process provides an increase in the average quarterly recovery (in dollars) from the correction of pharmaceutical price mistakes.

Recommendations

This work's area of focus is on the utility of an inductive methodology for determining pharmaceutical prices using an existing stream of invoice data from a large hospital system, yet this is not the end of the story. For one, there was no attempt to utilize a system which matched against known prices – that is, assembling and matching against a compendium of what prices *ought* to be. That style of approach is quite possibly a monumental undertaking considering the complexity of the marketplace and the multiple layers of distinction at every step (cost per drug per class of trade per account type per contract arrangement, etc.). Yet it is probable that the results of such a content-aware system can improve even upon the results of this model.

Researchers like Iacocca and Vallen (2021) have added context to such an approach with their paper concerning the prediction of pharmaceutical prices using a number of machine learning models, yet had faced limitations with regards to their access to real pricing data, and the less-than-favorable accuracy of their models. With a partnership with a larger healthcare provider and more discrete data about actual pricing, this methodology might yet yield larger and tastier fruit.

References

American Hospital Association. (2023). *Fact Sheet: The 340B Drug Pricing Program*. Retrieved April 29, 2023, from https://www.aha.org/system/files/media/file/2019/03/fact-sheet-340b-drug-pricinig-program-0119.pdf

American Hospital Association, Federation of American Hospitals, American Society of Health-System Pharmacists. (2019). *Recent Trends in Hospital Drug Spending and Manufacturer Shortages*. Retrieved April 22, 2023, from https://www.aha.org/system/files/2019-01/aha-drug-pricing-study-report-01152019.pdf.

Ang, S. (2014). *The research question*. SAGE Publications Ltd, https://doi.org/10.4135/9781473909694

Bahl, V., McCreadie, S. R., & Stevenson, J. G. (2007). Developing dashboards to measure and manage inpatient pharmacy costs. *American Journal of Health-System Pharmacy, 64*(17), 1859–1866. https://doi.org/10.2146/ajhp060596

Delgado, P. (2017, October 26). *Price disparity types. Identifying and correcting them*. Mirai. Retrieved April 19, 2023, from https://www.mirai.com/blog/price-disparity-types-identifying-and-correcting-them/

Elliott, A. C., & Woodward, W. A. (2016). *SAS Essentials: Mastering SAS for data analytics*. Wiley.

Health Resources & Services Administration. (2023, March). *340B Drug Pricing Program*. HRSA. Retrieved April 18, 2023, from https://www.hrsa.gov/opa

Healthcare Supply Chain Association. (n.d.). *What is a GPO?* HSCA. Retrieved April 25, 2023, from https://supplychainassociation.org/about-us/what-is-gpo/

Hisey, T., Jacoby, R., Heim, M., & Mancke, J. (2019). *The role of distributors in the US health care industry: 2019 report*. Deloitte. Retrieved April 20, 2023, from https://www2.deloitte.com/content/dam/Deloitte/us/Documents/life-sciences-health-care/us-hda-role-of-distributors-in-the-us-health-care-industry.pdf

Iacocca, K., & Vallen, B. (2021). Using Analytics to Gain Insights on U.S. Prescription Drug Prices: An Inductive Analysis. *Journal of Public Policy & Marketing, 40*(4), 538–557. https://doi.org/10.1177/0743915621993173

Jebb, A. T., Parrigon, S., & Woo, S. E. (2017). Exploratory data analysis as a foundation of inductive research. *Human Resource Management Review, 27*(2), 265–276. https://doi.org/10.1016/j.hrmr.2016.08.003

Karakostas, B. (2016). Event Prediction in an IoT Environment Using Naïve Bayesian Models. *Procedia Computer Science, 83*, 11–17. https://doi.org/10.1016/j.procs.2016.04.093

Kimball, R., & Ross, M. (2013). *The Data Warehouse toolkit: the Definitive Guide to Dimensional Modeling* (3rd ed.). Wiley.

Mulcahy, A. W., Whaley, C. M., Tebeka, M. G., Schwam, D., Edenfield, N., & Becerra-Ornelas, A. U. (2021). International Prescription Drug Price Comparisons: Current Empirical Estimates and Comparisons with Previous Studies. In *RAND Corporation eBooks*. https://doi.org/10.7249/rr2956

National Drug Code Directory. (2022). *U.S. Food And Drug Administration*. https://www.fda.gov/drugs/drug-approvals-and-databases/national-drug-code-directory

Parasrampuria, S. and Murphy, S. (2022). *Trends in Prescription Drug Spending, 2016-2021*. Washington, DC: Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. Retrieved April 21, 2023, from https://aspe.hhs.gov/sites/default/files/documents/88c547c976e915fc31fe2c6903ac0bc9/sdp-trends-prescription-drug-spending.pdf

Prescription Analytics, Inc. (2022). *Class of Trade* [White Paper]. Retrieved April 25, 2023, from https://prescriptionanalytics.com/wp-content/uploads/2022/12/Understanding_Class_of_Trade-COT-in_the_Pharmaceutical_Industry.pdf

RS & F Healthcare Advisors. (2019). Maryland Local Health Department – Billing Manual. In *Office of Population Health Improvement*. Maryland Department of Health. Retrieved April 29, 2023, from https://health.maryland.gov/pophealth/Documents/Local%20Health%20Department%20Billing%20Manual/PDF%20Manual/Section%20II/National%20Drug%20Code%20(NDC)%20Info%20and%20Guide.pdf

SAS Institute Inc. (2013). *SAS/STAT® 13.1 User's Guide* [E-book]. https://support.sas.com/documentation/onlinedoc/stat/131/anova.pdf

Shelley, S. (2022, August 11). *Pharma distribution: Carving New Ground*. Pharmaceutical Commerce. Retrieved April 23, 2023, from https://www.pharmaceuticalcommerce.com/view/pharma-distribution-carving-new-ground