

An Introduction to Healthcare Analytics

From Simple to Complex



Nashville Analytics Summit: Oct 18-19, 2021

Tools and Code Suggested For This Workshop

- MS Excel
- Open-source SQL (Postgres, My SQL, etc.)
- R/RStudio IDE and R file (Markdown, .R, etc.)
- Python IDE (I will use Jupyter Notebook) and .ipynb or .py file
- Data Files
- Your full and undivided attention

Jay Roy, CPA, MBA

- 20 years in corporate finance from Accountant to CFO
- 15 years Healthcare professional in Finance and IT
- 15 years as an entrepreneur – 3 startups and upcoming startup *
- Functional expertise as Product Manager / Project Manager / Financial / Healthcare Compliance
- 10 years Data and Analytics professional with SQL, R and Python programming
- Mentor / Educator / Speaker / * Entrepreneur – <https://nextgenhealthanalytics.com>

Jay Roy, CPA, MBA Continued

- ✚ 2019 Nashville Analytics Summit – “Leverage Your Customer Data and Look Like a Rockstar”
- ✚ 2020 Nashville Analytics Summit Judge
- ✚ 2021 IIBA Beyond Analysis Conference Speaker – “So You Want to be a Data Scientist, Kinda?”
- ✚ 2021 Lipscomb Masters of Healthcare Administration Speaker – “Introduction to Predictive Analytics”
- ✚ 2021 ACHE MT Leadership Program - “Using Data-Driven Decision Making to Make Your Team Successful”
- ✚ 2021 IIBA San Diego Chapter Speaker - “So You Want to be a Data Scientist, Kinda?”

Megan Heim

- 20 years in healthcare supporting performance improvement through technology and services
- 10 years + with two of the largest healthcare services companies, The Advisory Board and Optum
- Over 6 years with top-ranked leadership and recruitment firms that served the healthcare industry
- Deep understanding of the healthcare landscape and ability to assess the influences that impact organizations and their executive teams
- Experienced coach, mentor and recruiter and contributor to local and national healthcare organizations such as the ACHE, WITT and colleges.

Who we are ... <https://nextgenhealthanalytics.com>

[HOME](#)[ABOUT US](#)[COURSES](#)[REGISTER](#)[STUDENTS - LOG IN](#)

Welcome to

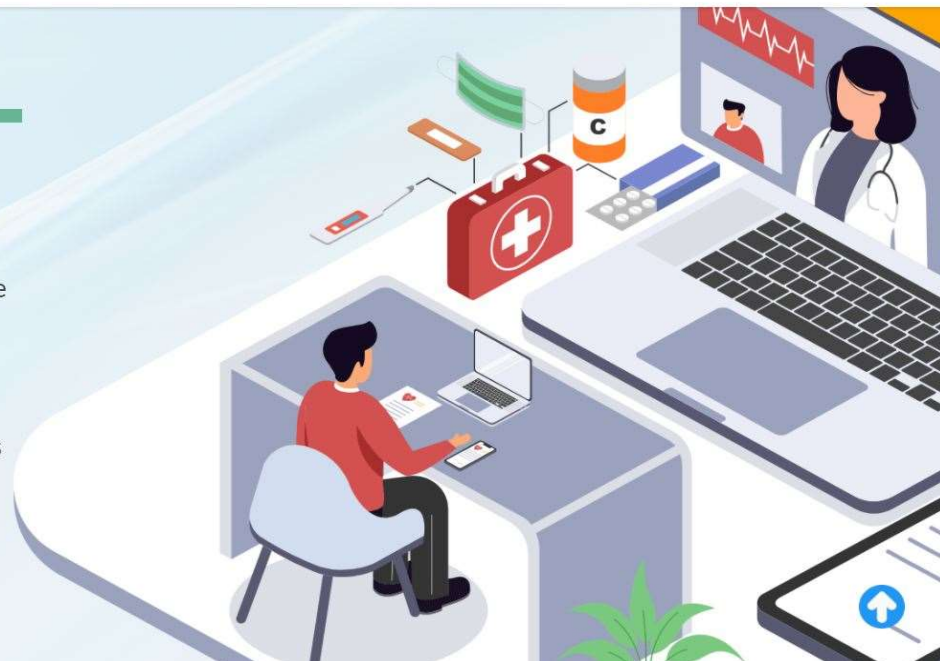
NGHA Analytics Academy

We are on a journey, like yourself, building the best course offerings, so you can be your best.

Our course offerings will prepare you to expand your knowledge and teach you how to approach and solve problems in an analytical way.

Our offerings provide you a combination of hard skills, soft skills and good advice from people "who have been there and done that" so you can benefit from their experiences.

This is one of the many unique differences compared to other course offerings you may be considering.



NGHA Mission ... One Simple Goal

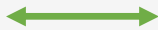
“ Creating the next generation of healthcare analytical professionals ”



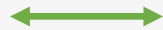
Agenda for Workshop



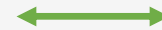
Healthcare Landscape
and Business Rules



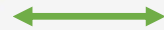
Triple Aim



Leadership

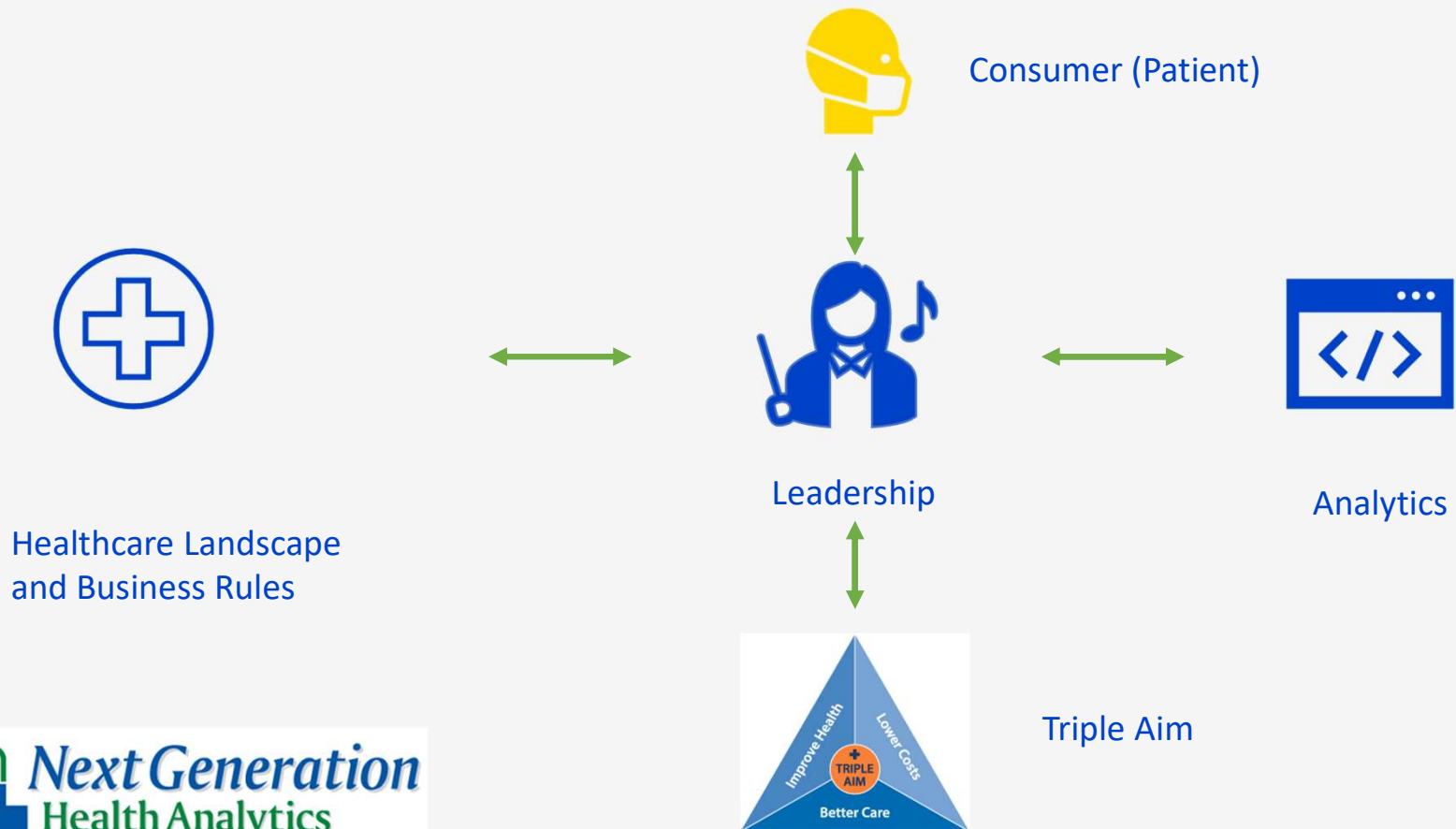


Consumers /
Patients



Analytics

Real World



Healthcare Landscape and Business Rules



- ✚ Healthcare Ecosystem
- ✚ Life of a Claim
- ✚ Revenue Models
- ✚ Types of Healthcare Data
- ✚ Business Rules
- ✚ Healthcare Data Challenges

Healthcare Ecosystem: Complex?

Medical Providers



Insurance Payers,
CMS/MACs



Government
Feds / State



Medical Institutions
e.g. (AMA)



Institute for
Healthcare
Improvement



World Health
Organization

Medical Research



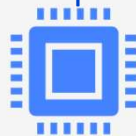
Software companies



Consumer (Patient)



Med-Tech
companies



Med-Devices
companies



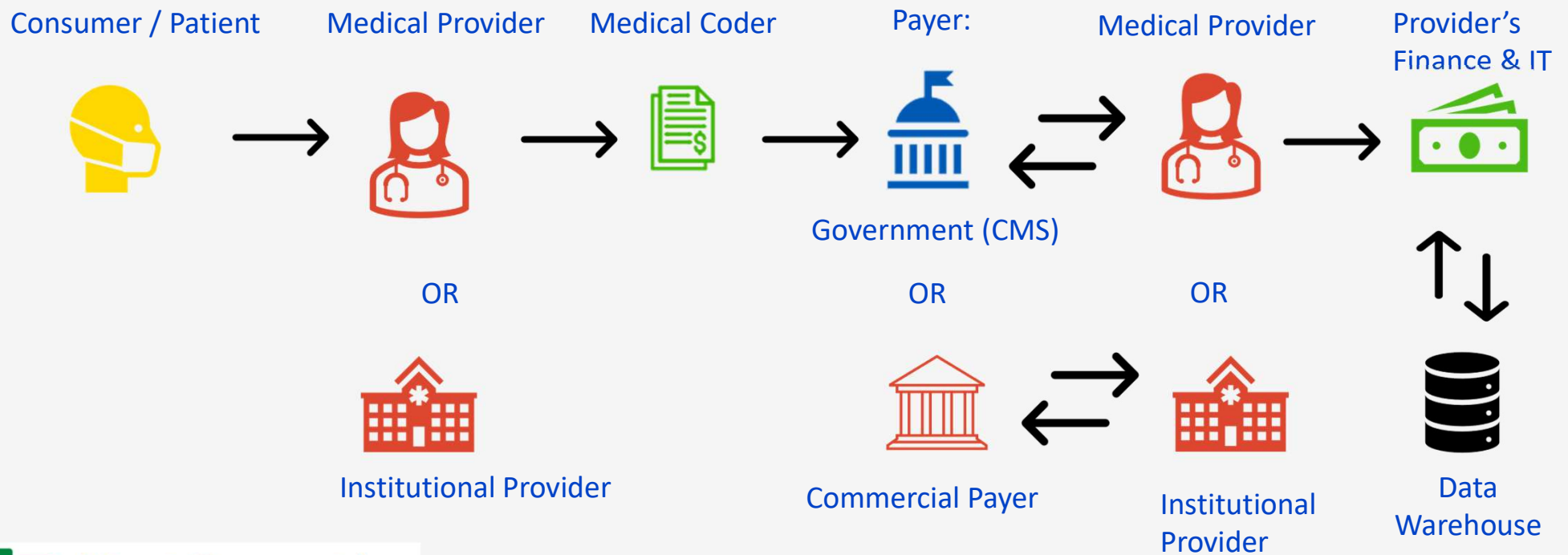
Pharma



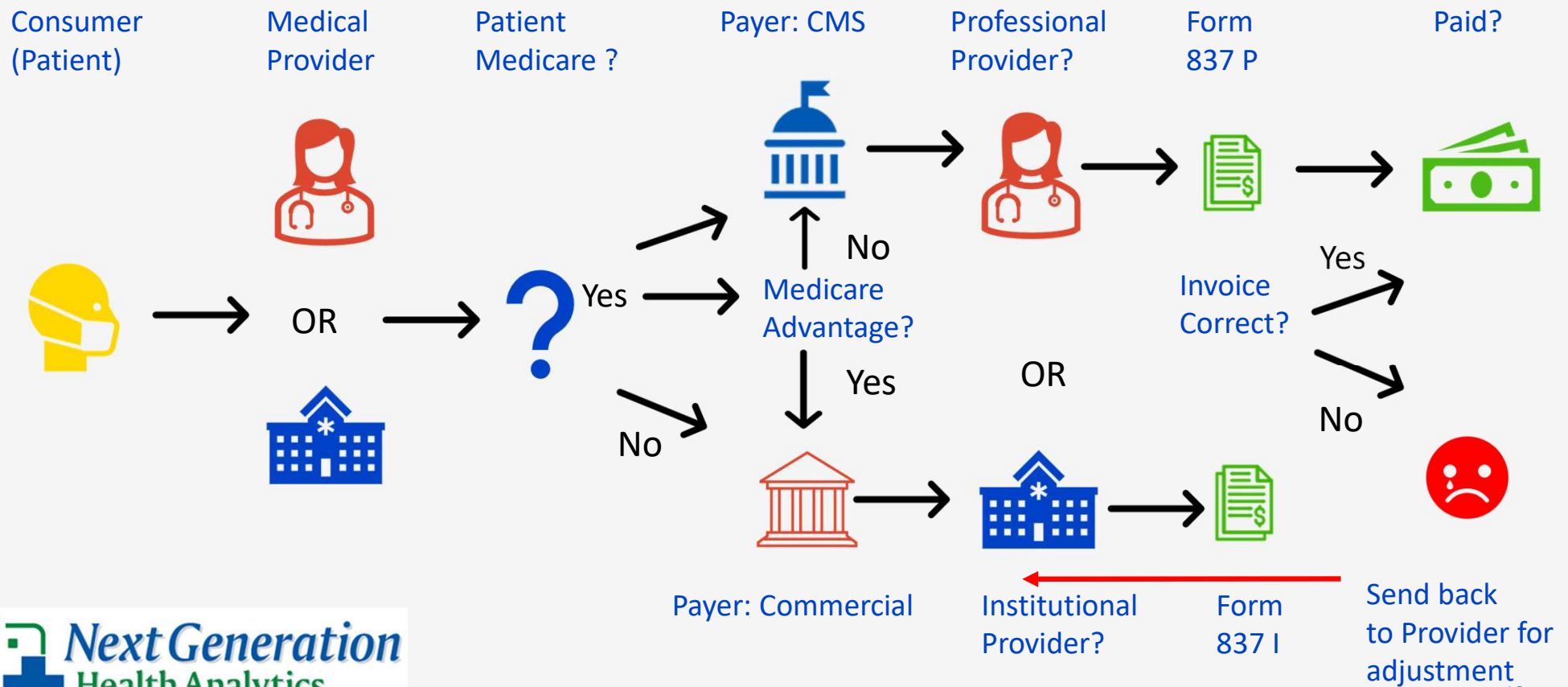
Retail &
Pharmacies



Life of a Claim - Revenue Cycle



Revenue Model: Traditional Fee-For-Service Charge



Revenue Model: Value-Based Care Charge

New Model

“Value-Based”



Type of Provider

- Traditional: Hospitals, etc.
- New: MSO / ACO / etc.



Alternative Payment Models

- Pay for Performance
- PCMH Payments
- Specialty Care
- Bundled Payments
- Shared Risk ACOs
- Full Risk ACOs



Types of Healthcare Data



PATIENT



PROVIDERS



PAYERS



QUALITY



REVENUE



ELIGIBILITY &
ENROLLMENT



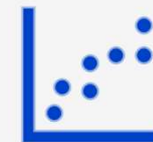
CLINICAL



CLAIMS



OPERATIONS



ANALYTICS

Healthcare Business Rules: Simple Example with SQL

- ✚ In healthcare, there are a myriad of business rules, dependencies and exceptions that are often unknown but important to consider since many of your analytics are predicated on them
- ✚ For example, using the Fee-For-Service revenue model, depending on the patient type, Medicare or not, a charge can be routed down different paths
- ✚ Continuing the example, if patient is Medicare, are they Medicare Part A, B , C and or D ?
- ✚ Considering the example above, without knowing the business rules your resulting SQL query may be incorrect since it is impacted by the WHERE clause and or CASE statements

Healthcare Business Rules – Type of Bill (“TOB”) FL 4

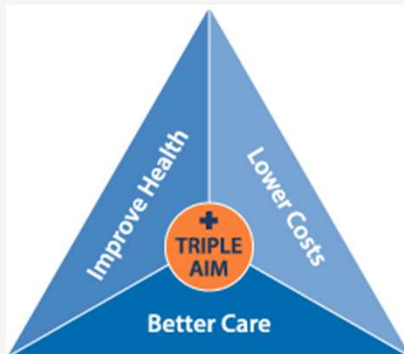
- ✚ Institutional Providers (Hospitals, etc.) must use CMS Form 1450 to bill for facility charges
- ✚ Within the Form 1450, the requirements define the business rules for each required field locator (“FL”) for the claim to be paid correctly by CMS. Specifically, item FL 04, (“Type of Bill”) , a 3 or 4-character number, must be entered in this field by the provider whether sent by paper or electronic sent (837I)
- ✚ Each of the 4/3 characters for FL04, has an individual meaning which identifies the type of charge to be paid by CMS. For example, the electronic format is 3 characters, 1st and 2nd character represents the type of facility, 1= Hospital, 2= Skilled Nursing Facility, 3rd position represents bill type, i.e. frequency of the bill
- ✚ Therefore, the Analyst could present inaccurate results due to lack of understanding the healthcare business rules, e.g. rules regarding correct CMS form, field locator/data field, and correct data for query

Healthcare Data Challenges Affect All Participants

- ✚ All participants (“3 Triple P*”) are affected by data quality issues but are not equally impacted
- ✚ Patients want to know what services cost, understandable and accurate bills and no surprises
- ✚ Payers want to know how much services cost and are costs preventable/avoidable
- ✚ Providers want to know patients are receiving quality care, cost efficient solutions, and great experience
- ✚ All parties want to do all the above but “Better, Faster, Cheaper”

* Payers, Providers, Patients

The Triple Aim ... Goal: Improve Healthcare



- ✓ Improve Patient Experience
- ✓ Reducing Cost per Capita
- ✓ Improve Health of Populations

<http://www.ihl.org/Engage/Initiatives/TripleAim/Pages/default.aspx>

Use Analytics & The Triple Aim to Improve Healthcare



- ✓ Improve Patient Experience
- ✓ Reducing Cost per Capita
- ✓ Improve Health of Populations



Achieving Improvements through Leadership and Diversity



Leadership



Diversity

Top 10 Qualities of a Great Leader

1. Vision
2. Inspiration
3. Strategic & Critical Thinking
4. Interpersonal Communication
5. Authenticity & Self-Awareness
6. Open-Mindedness & Creativity
7. Flexibility
8. Responsibility & Dependability
9. Patience & Tenacity
10. Continuous Improvement



Achieving Improvements through Leadership and Diversity

■ INFLUENCE is a critical skill – manage up and down

■ Set expectations. Be comfortable saying no.

■ Diversity, Diversity, Diversity

■ Leadership is a continuum

■ Who are you developing?

■ How do you motivate?

Leadership can mean many things to different people. An effective leader is "a person who does the following:

- Creates an inspiring vision of the future.
- Motivates and inspires people to engage with that vision.
- Manages delivery of the vision.
- Coaches and builds a team, so that it is more effective at achieving the vision."

How to develop the character traits of a leader

Prioritize learning and training

Ask for more responsibilities

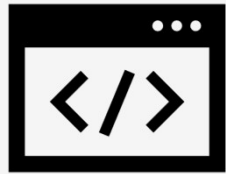
Improve your communication skills

Learn from a mentor

Adopt a Leadership style



Show and Tell ... Analytics



Case Study and Vignettes Using 4 Technologies

Case Scenario

- You are the new Chief Data Scientist at Better Health Healthcare System (“BHHS”)
- You are tasked by the CEO with using the Triple Aim Framework to improve the financial standing of BHHS while improving patient experience and lowering costs by identifying high-cost populations
- Working with the CIO of BHHS, after several positive conversations with her, it is important to improve the quality of data, business processes and educate team members of healthcare analytics
- Requirement: You are asked to demonstrate 4 analytics examples of challenges faced by BHHS so resources can be allocated to achieve the Triple Aim goals and improve profitability of BHHS

Case Study Analysis



Where to start?



Understand the Process Flow



Identify Point Solutions

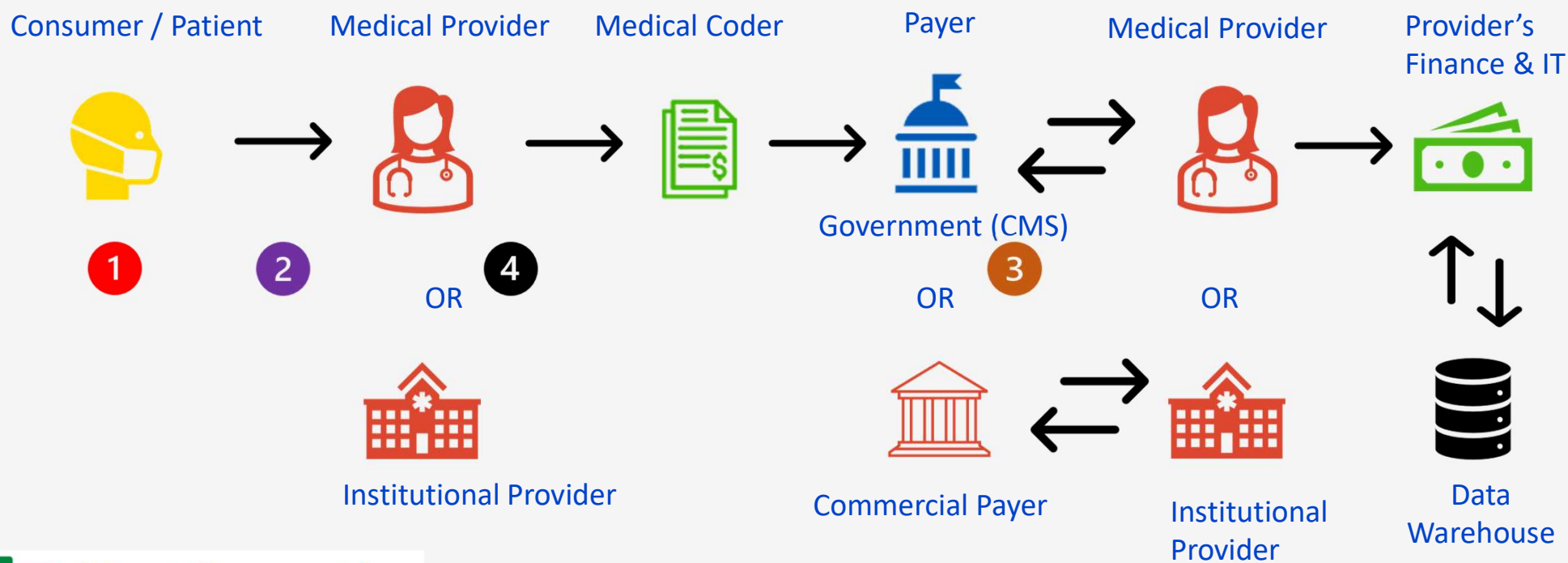


Use the Triple Aim as North Star

Case Scenario and 4 Analytic Vignettes

1. Patient identification example using Excel
2. Patient identification resulting from enrollment process using SQL
3. Predicting medical claims costs using R for Payers
4. “ The pathway ” to predicting expensive providers using Python

Examples of Data Quality Issues during the Payment Process



```
graph TD; HD[Historical Data] --> MB[Model Building]; MB --> ME[Model Evaluation]; ME --> MO[Model Optimization]; MO --> MB; ME --> ND[New Data]; ND --> AN[Answers]; MO --> AN; AN --> CR[Communicate Results]; CR --> DT[Data Transformation]; DT --> MB; CR --> Ex4[Example 4];
```

The diagram illustrates the Machine Learning process, organized into four numbered steps:

- 1 Examples**: Historical Data
- 2 Modeling**: Model Building, Model Evaluation, Model Optimization
- 3 Prediction**: New Data, Answers
- 4 Example**: Communicate Results

The flow is as follows:

- Historical Data (Step 1) feeds into Model Building (Step 2).
- Model Building (Step 2) feeds into Model Evaluation (Step 2) and Model Optimization (Step 2).
- Model Evaluation (Step 2) feeds into New Data (Step 3) and Answers (Step 3).
- Model Optimization (Step 2) feeds into Answers (Step 3).
- New Data (Step 3) and Answers (Step 3) feed into Communicate Results (Step 4).
- Communicate Results (Step 4) feeds back into Data Transformation (Step 4), which then feeds back into Model Building (Step 2).

Scenario 1 – Patient Identification Using Excel

- Objective : Can we correctly identify Patients ?
- Challenge 1 : Different stakeholders identify patients/beneficiaries by different identifiers which can cause confusion within provider organizations and across payers, providers and government
- Impact: For all healthcare ecosystem parties, an ongoing battle for ensuring data quality which needs many checks and balances from both an operational and data perspective. Otherwise, GIGO.

Scenario 1 – Patient Identification Provider View

Patient Table

pat_id	pat_name	add_line_1	zip	birth_date	Age	sex	ssn	reg_date	reg_status_c	medicare_num	medicaid_num	mother_pat_id	father_pat_id
Z0000001	PENN,GEORE	1553 UNIONPORT RD	7644	12/25/2001	19.8	M	000-00-0000	10/12/2010	2	NULL	NULL	NULL	NULL
Z0000002	MEN,SHARONE	1155 EAST 225TH STREET	7644	6/3/1993	28.3	F	000-64-0000	1/1/1900	3	NULL	NULL	NULL	NULL
Z0000003	MARTIN,JESS	828 WARING AVE	7643	5/5/1969	52.4	F	000-60-0000	3/24/2011	2	NULL	NULL	NULL	NULL
Z0000004	MCREA,REM	1095 UNIVERSITY AVENUE	7605	1/26/1998	23.7	F	000-00-0099	4/18/2015	2	NULL	NULL	NULL	NULL
Z0000005	SANTIAGO, YO	3045 VILLA AVE	7676	2/26/1974	47.6	F	000-00-0011	7/26/2010	2	NULL	NULL	NULL	NULL
Z0000006	SNEEZE,DAISY	500 Union ave	10455	7/4/1976	45.2	F	000-00-0012	4/30/2019	2	NULL	NULL	NULL	NULL
Z0000007	JIMMY,RALPH	2475 SOUTHERN BOULEVARD	10987	2/28/1967	54.6	M	000-00-0013	10/15/2017	2	NULL	NULL	NULL	NULL
Z0000008	SANTA,CLAUSE	1615 NORTH POLE	10876	7/4/1965	56.3	M	000-00-0014	2/2/2015	2	NULL	NULL	NULL	NULL
Z0000009	BEAR,TRISH	4289 HOLIDAY AVE	10567	12/25/1997	23.8	M	000-00-0015	6/24/2011	2	NULL	NULL	NULL	NULL
Z0000010	ROUNDER,CORNER	115 INTER AVE	10678	8/5/1996	25.1	F	000-00-0016	8/17/2010	2	NULL	NULL	NULL	NULL
Z0000011	PANCH,SUSAN M	2011 ATLAS PLACE	10345	4/1/1973	48.5	F	000-00-0017	4/25/2014	2	NULL	NULL	NULL	NULL
Z0000012	SMITH SR., ROBERT	105 ARMISTEAD DR.	75077	4/1/1954	67.5	M	000-00-0018			1EG5SX5MK73		NULL	NULL
Z0000013	SMITH, JOANNE	105 ARMISTEAD DR.	75077	8/12/1955	66.2	F	000-00-0019			1EH5FE6NL14		NULL	NULL
Z0000014	SMITH JR., ROBERT	105 ARMISTEAD DR.	75077	4/1/1979	42.5	M	000-00-0020			NULL		Z0000013	Z0000012
Z0000015	SMITH, ROBERTA	105 ARMISTEAD DR.	75077	4/1/1975	46.5	F	000-00-0021			2EG5SX5MK74		Z0000013	Z0000012

Scenario 1 – Patient Identification Payer View

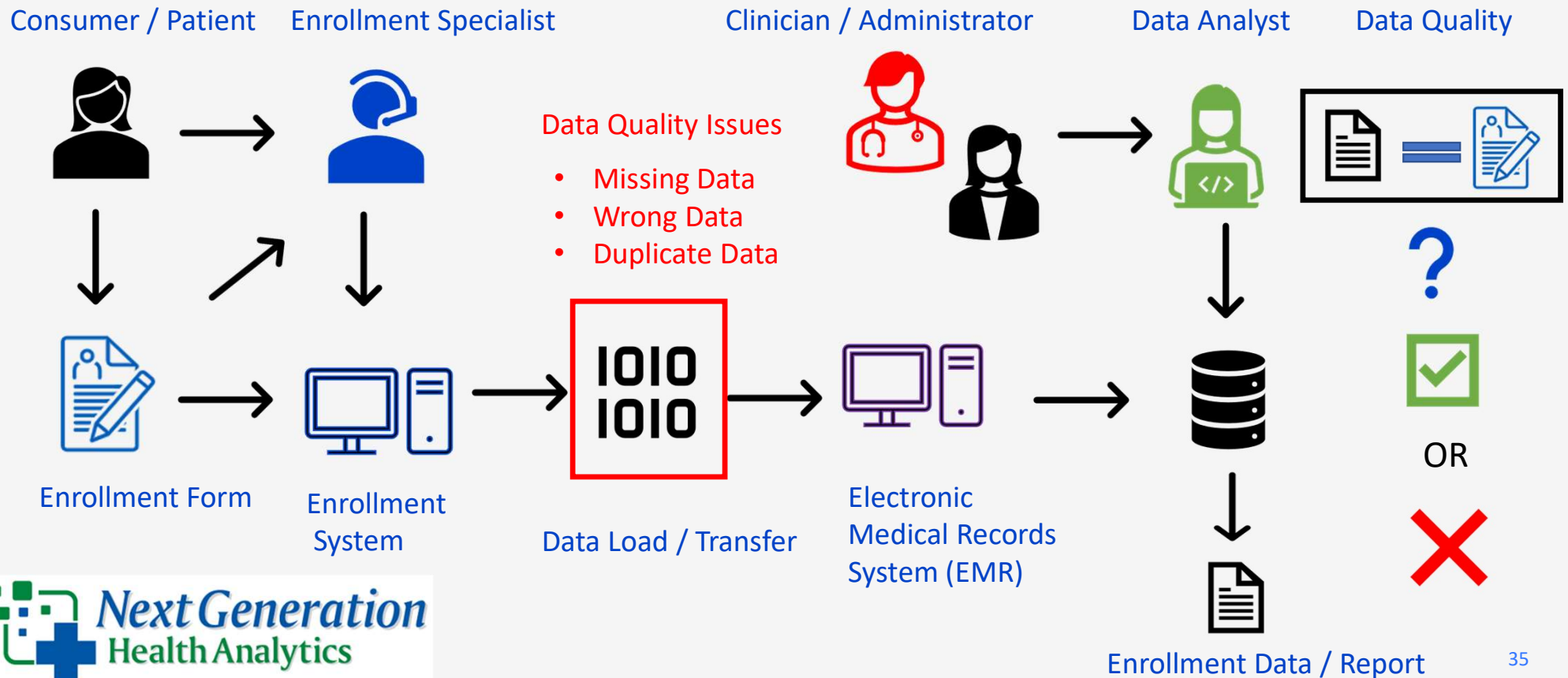
Coverage Table

subscriber_id	member_id	mem_rel	dob	first_name	middle_name	last_name	address	medicare_number_MBI	group_name	group_number	part_d_gro
S000000000000001	M0000000000000001	18	02-22-1955	Buddy	Bear	Retreiver	145 Main St.	NULL	roydogs001	1407	NULL
S000000000000002	M0000000000000002	14	02-22-1955	Diamond	Mojo	Dobe	3114 Calstone	NULL	DOGON	1409	NULL
S000000000000003	M0000000000000003	18	02-22-2020	Casey	Cadodoole	Labrador	3114 Calstone	1EG4-TE5-MK73	mcr_1	NULL	
S000000000000001	M0000000000000004	1	02-22-1962	Maggie	Mooster	Retreiver	145 Main St.	NULL	roydogs001	1407	NULL
S000000000000004	M0000000000000005	18	09-11-2020	Leo	Roy	Maincoon	100 BeltLine Rd.	2EG4-TE5-MK74	mcr_1	NULL	
S000000000000005	M0000000000000006	18	02-22-2002	Charlie	Tuna	Tux	2499 Long Prairie	NULL	73120	170002	NULL
S000000000000006	M0000000000000007	18	04-01-1954	ROBERT		SMITH SR.	105 ARMISTEAD DR.	1EG5SX5MK73	CofNash	10225	
S000000000000006	M0000000000000008	2	04-01-1979	ROBERT		SMITH SR.	105 ARMISTEAD DR.	NULL	CofNash	10225	

Scenario 2 – Patient Identification Enrollment Using SQL

- Objective : Can we correctly identify Patients – Part II (Eligibility and Enrollment)
- Challenge : The Enrollment Intake process is completed without good quality resulting in downstream analytics such as member and patient counts to be inaccurate
- Impact: The importance of the Enrollment process should not be underestimated because many of the downstream processes rely on the quality at the front-end (enrollment) such as paying claims correctly, aggregated data making sense and preventing downstream analytics from impairment

Enrollment Process and Populating Database Tables



Scenario 2 – Patient Identification Data Quality


Coverage Table -

Data Output	Explain	Messages	Notifications
subscr_name character varying (160)	payor_id numeric (18)	plan_id numeric (18)	subscr_num character varying (50)
1 MARTIN,JESS	3002	300201	ZC99065P
2 MARTIN,JESS	2001	900801	4567890
3 MARTIN,JESS	3002	501301	IKS89076P04
4 MARTIN,JESS	3002	197500	IKS89076P04
5 MARTIN,JESS	107	901001	987867564

Note:

Three possible answers based on different identifiers (1. name, 2. ssn, 3. pat_id

Data Output	Explain	Messages	Notifications	
 pat_rec_of_subs_id character varying (18)	 subscr_name character varying (160)	 payor_id numeric (18)	 plan_id numeric (18)	 subscr_num character varying (50)
1 Z0000003	MARTIN,JESS MARGARET	2001	200102	134506789B
2 Z0000003	MARTIN,JESS	3002	300201	ZC99065P
3 Z0000003	MARTIN,JESS	2001	900801	4567890
4 Z0000003	MARTIN,JESS	3002	501301	IKS89076P04
5 Z0000003	MARTIN,JESS M	3002	900901	IKS89076P04
6 Z0000003	MARTIN,JESS	3002	197500	IKS89076P04
7 Z0000003	MARTIN,JESS M	3002	197500	ZC99065P
8 Z0000003	MARTIN,JESS M.	3002	501001	ZC99065P
9 Z0000003	MARTIN,JESS	107	901001	987867564

Data Output	Explain	Messages	Notifications		
 pat_rec_of_subs_id character varying (18)	 subscr_name character varying (160)	 payor_id numeric (18)	 plan_id numeric (18)	 subscr_num character varying (50)	
1	Z0000003	MARTIN,JESS MARGARET	2001	200102	134506789B
2	Z0000003	MARTIN,JESS	3002	300201	ZC99065P
3	Z0000003	MARTIN,JESS	2001	900801	4567890
4	Z0000003	MARTIN,JESS	3002	501301	IKS89076P04
5	Z0000003	MARTIN,JESS M	3002	900901	IKS89076P04
6	Z0000003	MARTIN,JESS	3002	197500	IKS89076P04
7	Z0000003	MARTIN,JESS M	3002	197500	ZC99065P
8	Z0000003	MARTIN,JESS M.	3002	501001	ZC99065P
9	Z0000003	MARTIN,JESS	107	901001	987867564

Scenario 3 – Predicting Medical Claims Using R for Payers

- Objective : Consistently predicting costs is both art and science filled with pitfalls
- Challenge : Your teams need to understand your business processes, data and drivers or else analytical processes can't be scaled and be agile
- Impact: Using the Triple Aim tenets of lowering costs, understanding the drivers of cost(s) and improving patient population health in detail is paramount
- Context: BHHS has a business unit (medical insurance) for paying medical claims

Scenario 3 – Predicting Medical Claims Using R for Payers

- Situation : The dataset is a small with 7 fields and ≈ 1300 rows. The dataset will need some level of feature engineering including demographic predictors like smoker status
- The patient population are Non-Medicare beneficiaries, between the ages of 18 – 64
- Since we are dealing with predicting medical costs, dependent variable ($Y = \text{Expenses}$), linear model will be used for predictions. The R package(s) used are psych and stats
- Hypothesis : Do demographic predictors of age, smoker, & bmi drive medical costs ?

Scenario 4 – Pathway to Prediction Expensive Providers

- Objective : How expensive are Providers at delivering services and managing costs ?
- Challenge: Developing and testing prediction algorithms, highlights the degree of manual and flawed business processes impacting data
- Context: BHHS serves older patient population, the dataset comes from CMS and billable codes are HCPCS and the providers are listed on NPPES registry (<https://npiregistry.cms.hhs.gov/>)

Scenario 4 – Pathway to Prediction Expensive Providers

- The dataset is comprised of 2019 data with 73 fields and ≈ 1 Million rows. It will need some level of feature engineering including dealing with missing data, non-relevant predictors, etc.
- Feature engineering includes dealing with missing data, non-relevant predictors, etc.
- We are dealing with predicting medical costs, dependent variable ($Y = \text{Tot_Mdcr_Pyamt_Amt}$), linear model would be used for predictions. Options for modeling can be various python packages, including: statsmodels, sklearn.linear_model.Linear Regression()

Scenario 4 – Predicting Expensive Providers – Cont’d

- Data profiling, exploration and assumptions of the model/data need to be considered
- Within healthcare data, there are LOTS of superfluous variables (predictors) that need to be considered and removed. This is the requirement and the art of healthcare analytics
- Hypothesis : Basic level, the type of provider and the services provided drives medical costs
- Consider multicollinearity, much of the data is correlated so profiling is VERY important
- Impact: It's a journey and requires patience since much of the “data work “ is due to the complexity of the ecosystem and dependencies between organizations

Summary and Insights

- ✚ Healthcare business rules are complex and sometimes quirky
- ✚ Unique feature engineering and data quality challenges
- ✚ Identifying the “best analytics” comes from the Triple Aim
- ✚ Communicating with healthcare professionals is both an art and science
- ✚ Leadership is required to take organizations from good to great

My Contact Information:

Jay Roy, CPA, MBA

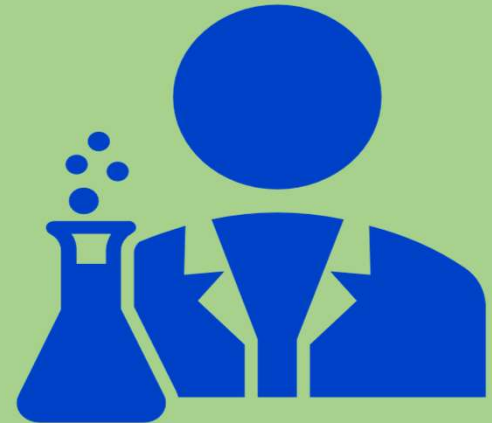
jay.roy@nextgenhealthanalytics.com

<https://linkedin.com/company/next-generation-health-analytics>

M: 931-919-8767

GitHub:

<https://github.com/jayroy1/NAS-2021-/>



-
- CIO / CFO / CPO
 - Product Data Scientist
 - Health Tech Entrepreneur
 - Educator / Mentor



An Introduction to Healthcare Analytics

From Simple to Complex



Nashville Analytics Summit: Oct 18-19, 2021