

Twin Cities Area SAS Users Conference

A hiker is walking on a rocky, volcanic trail. The landscape is rugged and colorful, with red and orange volcanic rock formations. In the background, there are mountains with patches of green and blue, possibly snow or mineral deposits. The sky is overcast with grey clouds. The hiker is wearing a light-colored jacket and a hat, and is carrying a backpack. The overall scene is a dramatic, natural landscape.

Episode Grouping and Predictive Modeling with Health Care claims

**Steve Wickstrom, Vice President, Research and
Methods, Health Management Solutions, Ingenix**

April 30, 2009

INGENIX®

Agenda

- **Organizational Overview**
- **Data**
- **Analytic Solutions**
 - **Episodes of Care**
 - **Predictive Modeling**
- **Discussion**

Overview

- **Ingenix**

- Subsidiary of UnitedHealth Group founded in 1996

- **Health Management Solutions**

- Develops health care solutions targeted primarily towards the payer (health plan) and employer market. The five main areas are:
 - Health Measurement Solutions
 - Care Management
 - Underwriting
 - Provider Data
 - Network Intelligence

- **Research and Methods**

- Provides research and business analytic support for all HMS solutions
- SAS is the primary analytic tool

Health Plan Administrative Data

- **This data is based up activities that occur with the health care system – typically within a health plans and includes, but is not limited to the following:**
 - **Transaction Data**
 - **Enrollment Data**
 - **Provider Data**
 - **Additional Data**

Transaction Data

- **Data submitted by the providers of health care for reimbursement from the health plan**
 - **Medical Claims (Professional and Ancillary)**
 - When: Date of Service
 - What: Procedure Code
 - Why: Diagnostic code(s)
 - Who for: Member
 - Who by: Provider
 - **Facility Claims (Inpatient and Outpatient)**
 - When: Date(s) of Service
 - What: Procedure Code, Revenue Code
 - Who for: Member
 - Who by: Facility

Transaction Data (cont)

- **Data submitted by the providers of health care for reimbursement from the health plan**

- **Pharmacy Claims**

- When: Date of fill
- What: NDC code
- How much: Days supply, count
- Who for: Member
- Who from: Subscribing Provider
- Who where: Retail Pharmacy

Transaction Data (cont)

- **Enrollment Data**

- Indicates when “member” is eligible for health plan services (and what services – e.g. benefits)
- Demographic data

- **Provider Data**

- Type of provider: Facility or provider specialty
- Location(s)

- **Other Data**

- Laboratory results data
- Health Risk Assessment (HRA)
- Consumer
- . . .

Transaction Data - Example

Single member's electronic medical records:

1 st DOS	Last DOS	Proc Code	1 st DX	2 nd DX ...	NDC	Comments
10-13-07	10-13-07	99213	38200	4658		Office visit for Otitis Media and Acute URI
10-13-07	10-13-07				00093415573	Pharmaceutical
3-12-08	3-12-08				00173045301	Pharmaceutical
3-12-08	3-12-08				00172439018	Pharmaceutical
3-12-08	3-12-08	99213	4619	49390		Office visit: Acute Sinusitis Asthma
3-12-08	3-12-08				0069312019	Pharmaceutical
3-12-08	3-12-08	99213	78650			Office visit for chest pain
5-27-08	5-27-08	99213	38200	4658		Office visit for Otitis Media and Acute URI
5-27-08	5-27-08				00093417774	Pharmaceutical
11-03-08	11-03-08				24208075006	Pharmaceutical

Putting it all together . . .

- **Data warehouse contains nationally representative transactional data for over 50 million individuals over 10 years of time**
 - **Data processed in a consistent way**
- **This data is used as a Research and Development “sandbox” as we develop our solutions that we bring to market**
- **Today we will focus on some of the analytical engine solutions that we have developed**

Symmetry Analytical Engines

- **Flat file in – flat file out**
 - The Symmetry solutions take in flat files and apply value added metrics populated in a resultant flat file
- **User designs analytics or creates applications for their specific business use**
 - User can select “unit of analysis” which can range from within a person (e.g. disease) to large populations
- **Covers four main areas of measurement**
 - Cost
 - Utilization
 - Quality of care
 - Risk

Symmetry Analytical Engines

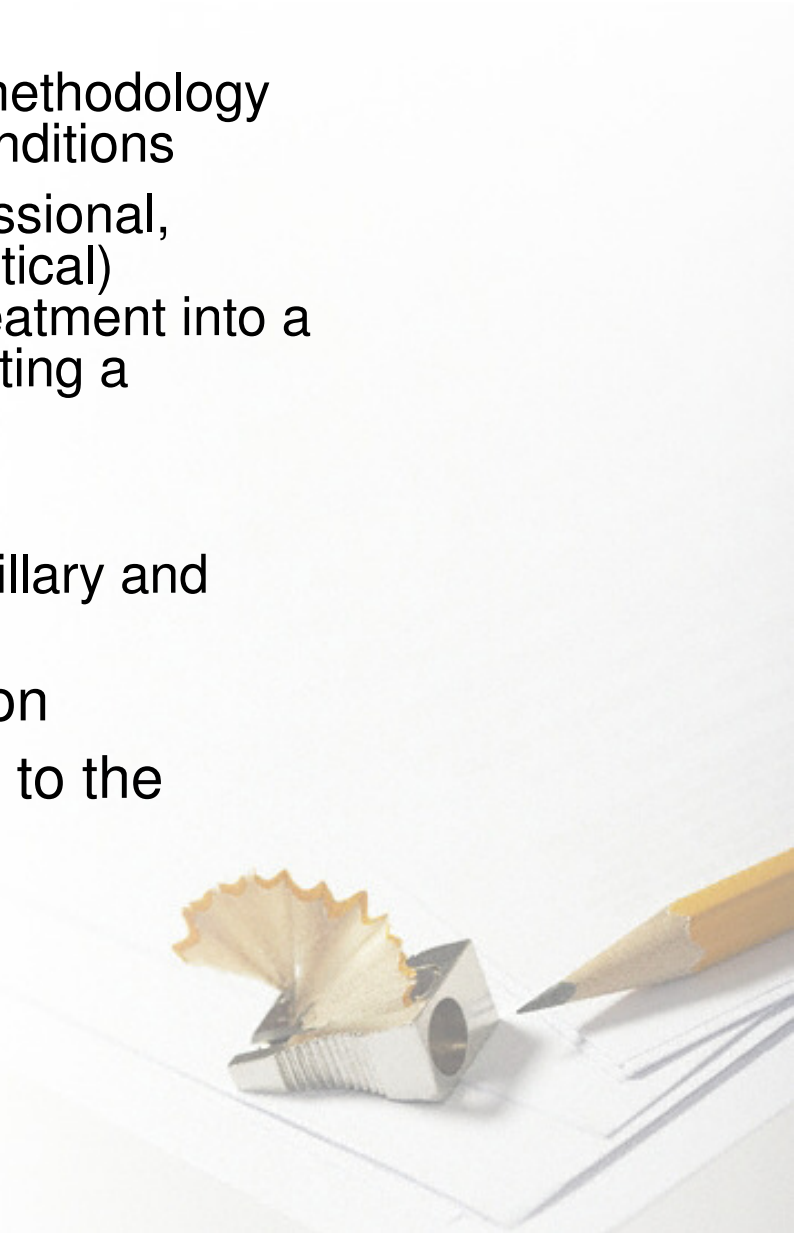
- **Episode Treatment Groups®**
 - Create episodes of care allowing diseases to be measured at the person level (cost and utilization)
- **Episode Risk Groups™**
 - A predictive modeling tool that measures individual risk using ETG as the basis of measuring morbidity
- **Pharmacy Risk Groups™**
 - Predictive modeling tool measuring individual risk using pharmacy data only
- **EBM Connect®**
 - Measures quality of care for selected conditions used evidence based medicine guidelines
- **Procedure Episode Groups™**
 - Measures services associated with specific therapeutic procedures

Research and Development Process

- **In order to create our analytic engine solutions we use SAS to leverage our large database to create the analytics/methodology behind our products:**
 - Descriptive statistics: Understand the problem and propose solutions
 - Complex programming to create metrics at the correct unit of analysis
 - Statistical analysis when appropriate - particularly useful for assessing risk
 - Creating a prototype that can be used for validation against the developed solutions

Example: Symmetry ETG®

- Symmetry ETG is an illness classification methodology that can be used to identify and quantify conditions
- Organizes relevant medical services (professional, inpatient, outpatient, ancillary & pharmaceutical) provided during the course of a patient's treatment into a medically relevant unit of analysis representing a patient's complete episode of care
- Claims Data Used:
 - Inpatient, Outpatient, Professional, Ancillary and Pharmaceutical
- Unit of analysis is condition within person
- Allows for analysis at any level relevant to the business
 - Case management
 - Disease management
 - Health plan assessment



ETG Concepts

- **Episode**

- A timeframe (occurrence) where all care for a condition is represented:
 - Professional, Facility or ancillary claims

- **Anchors & Non-Anchors**

- Anchor Records represent direct evaluation and/or management, or treatment of a patient (e.g. office visit, surgery)
- Non-Anchors represent care that is incidental to direct evaluation and/or management of a patient (e.g. laboratory, radiology or pharmacy)
- Each claim line is identified as either an Anchor or a Non-Anchor record

- **Clusters**

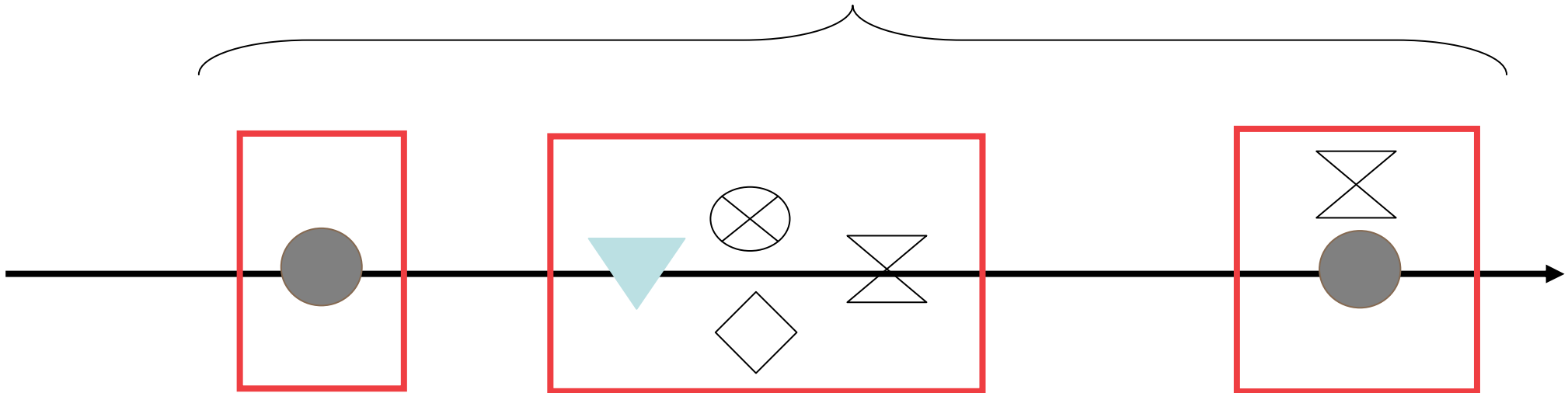
- Episodes are composed of Clusters of clinically similar information which include both Anchor and Non-Anchor records

- **Clean Periods**

- The time we look forward or backwards to determine if this is a 'new' instance of the conditions
 - Short for acute conditions – long for chronic conditions

ETG Concepts

There were three clusters of claims records for this particular episode of care for a particular condition.



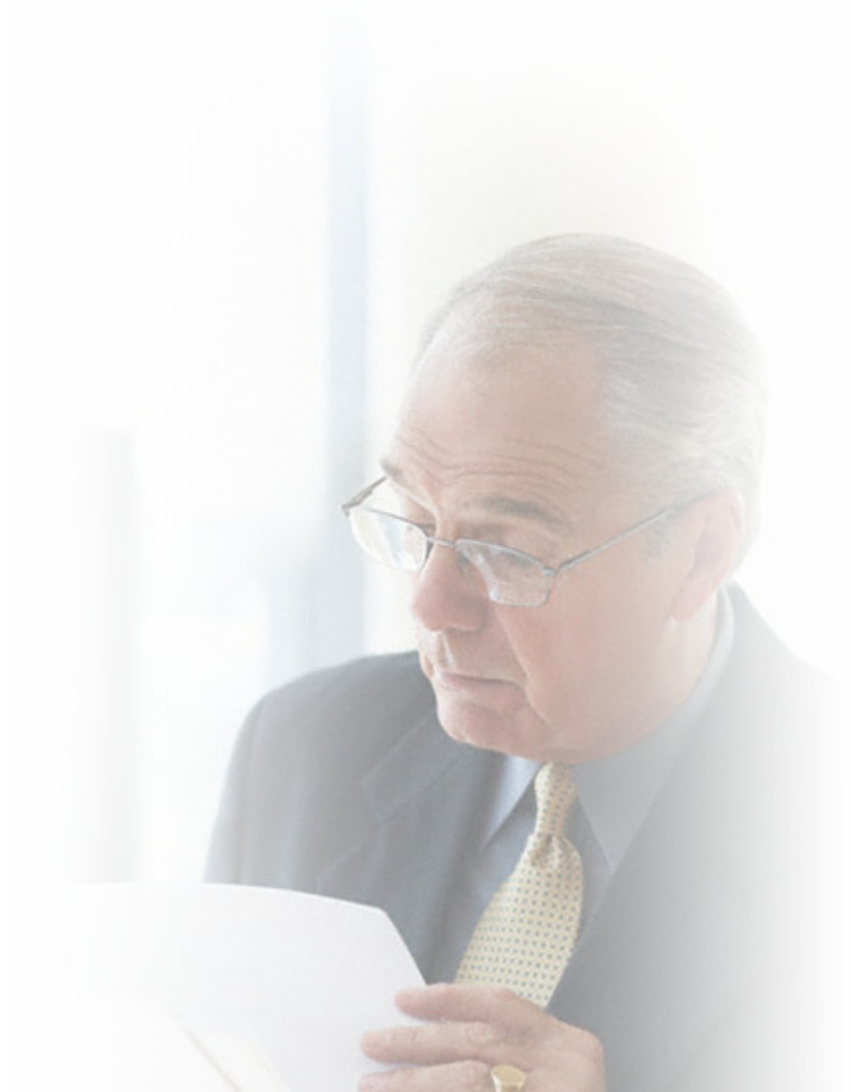
Key:

- Office Visit
- ▼ Inpatient Stay
- ⊗ X-ray
- ◇ Lab
- ⊠ Pharmaceutical

ETG Concepts

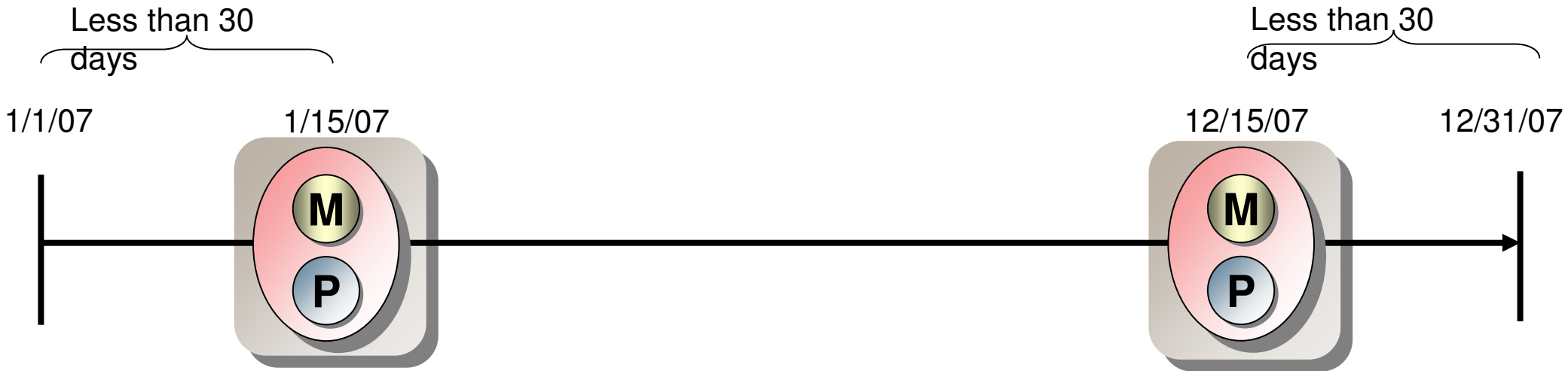
Clean Periods

- Each Symmetry ETG has its own unique clean period.
- Allows identification of multiple instances of the same acute condition.
- Each new Anchor record for that Symmetry ETG that falls within the “clean period” window will extend the length of the episode.

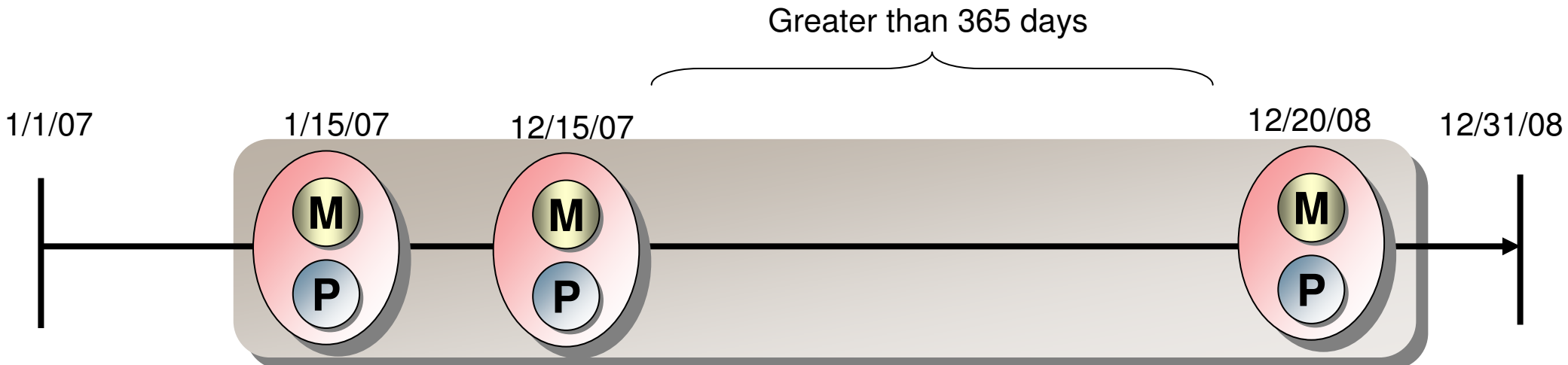


Episode Start/Finish

Acute Bronchitis – clean period 30 days



Diabetes – clean period 365 days



Building An Episode

- Assign Record Types
- Assign ETG to Anchor Records
- Placement of Signs & Symptoms Codes
- Placement of Non – Anchor Records
- Episode Types Are Assigned
- Assigning Complications, Comorbidities & Treatment Indicators
- Determination of Severity Level within condition
- Very complex programming to create a prototype – lots of details to track – with lots of changing and dependencies



Building An Episode

Single member's electronic medical records:

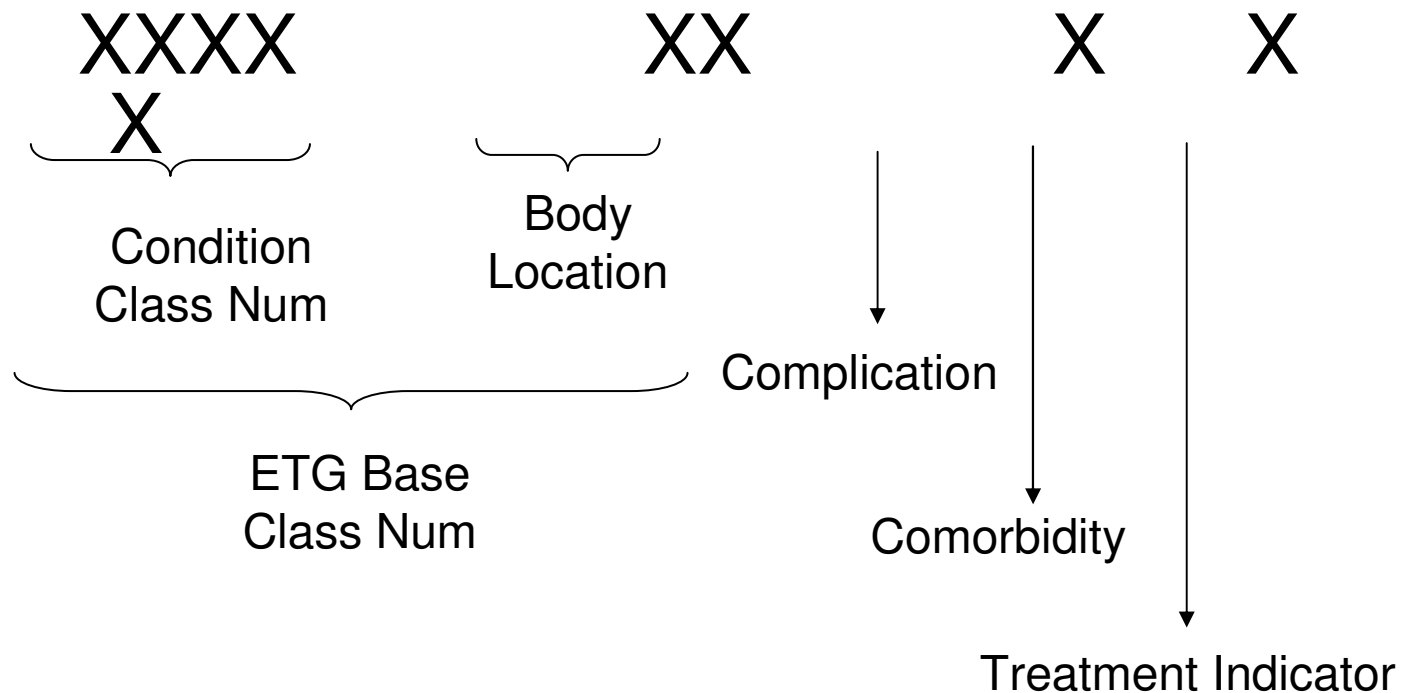
1 st DOS	Last DOS	Proc Code	1 st DX	2 nd DX ...	NDC	Rec Type	Condition	Episode
10-13-03	10-13-03	99213	38200	4658		M	402900	1
10-13-03	10-13-03				00093415573	P	402900	1
3-12-04	3-12-04				00173045301	P	403300	3
3-12-04	3-12-04				00172439018	P	438800	2
3-12-04	3-12-04	99213	4619	49390		M	438800	2
3-12-04	3-12-04				0069312019	P	403300	3
3-12-04	3-12-04	99213	78650			M	438800	2
5-27-04	5-27-04	99213	38200	4658		M	402900	4
5-27-04	5-27-04				00093417774	P	402900	4
11-03-04	11-03-04				24208075006	P		

Episodes: 1,4 = Otitis Media; 2 = Asthma; 3 = Acute Sinusitis

The End Result: Symmetry ETG Number

The Symmetry ETG Number is very rich in detail

– Nine (9) digit number



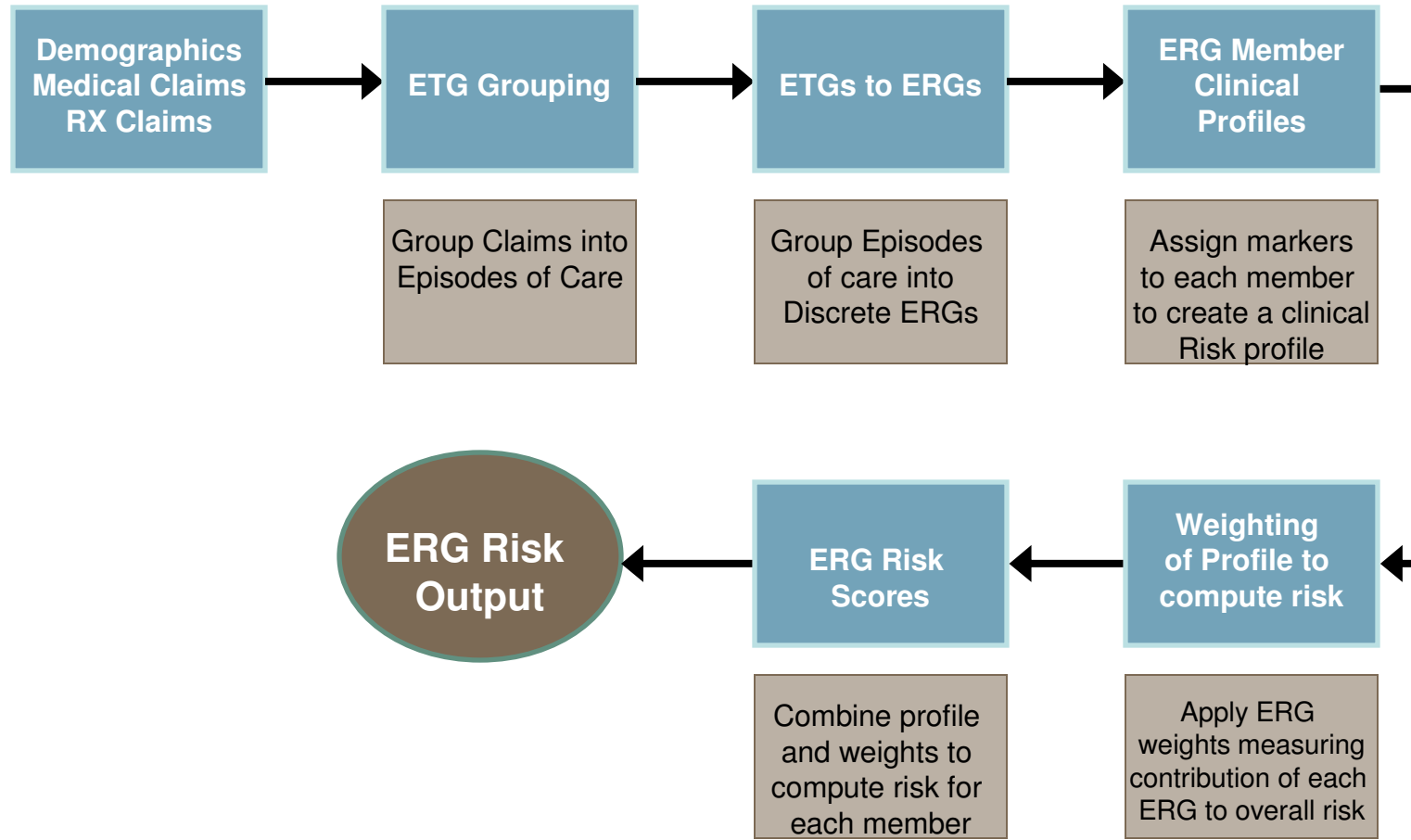
Episode Risk Groups™

- **Symmetry Episode Risk Groups™** create measures of current and future health risk for individuals and groups using medical and pharmacy claims information along with demographic data
- ERG uses episodes defined by the ETG™ grouper (along with age and gender) to assign relative risk scores to each member.
- Each individual risk score predicts a member's current (*retrospective*) and future (*prospective*) need for health care services and associated costs
- A relative risk score of 1.0 indicates average risk, above 1.0 indicates higher than average risk, and below 1.0 indicates lower than average risk.

ERG Overview

- **Risk assessment tool based upon ETGs**
 - Leverages ETG base conditions, severity levels
- **Product provides member level relative risk scores for multiple applications:**
 - Prospective: Care Management or Prospective Payment
 - Retrospective: Risk adjustment
 - Actuarial/Underwriting: Pricing of premiums
- **The Research team is tasked with providing the analytics for all of these**
 - As an example, we'll walk through what we needed to do in order to provide the content for our most recent release (ERG 7.0)

ERG Process



ERG Models

Features Assigned to Research Team

- Timing: Prospective, Retrospective, A/U
- Scores Provided on Output: Pro, Retro, A/U, Demo
- Input/Output: Med/Med, Med/MedRx, MedRx/MedRx
- Expenditure Thresholds: \$25,000, \$100,000
- Partial Enrollment: 1-3 months, 4-6, 7-9, 10-12
- We need to create analytics to support all of these

Where are we heading?

Assessment of Risk at the Individual Level

Example: Male age 58 Model: MedRx-MedRx, 100k Threshold, 10-12 Months								
ETG	Sev Lev	Desc	ERG	Description	Retro Risk	Pro Risk	A/U Risk	Demo Risk
163000	2	Diabetes	02.022	Diabetes with significant complication/ comorbidity, I	0.9874	1.2810	1.3175	---
386800	1	Congestive Heart Failure	08.043	Ischemic Heart Disease, heart failure, cardiomyopathy, III	2.2870	2.0065	1.9978	---
473800	3	Ulcer	11.022	Other moderate cost gastroenteology, II	2.3972	0.6474	0.5300	---
666700	1	Acne	17.011	Lower cost dermatology, I	0.1409	0.1023	0.1055	---
666700	1	Acne	17.011	Lower cost dermatology, I	---	---	---	---
Demographic Risk Group: Males, 55-64					n/a	0.7331	0.8229	2.0935
Total Risk Score					5.8125	4.7703	4.7737	2.0935

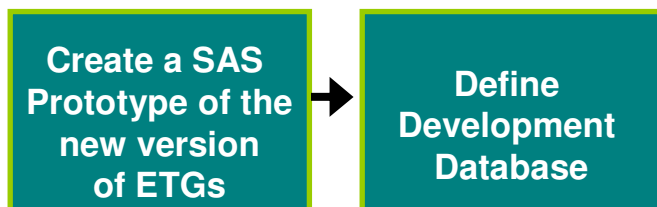
We need to develop the analytics to create all of the risk weights to support the above computations.

Development Methodology – Step 1

Create a SAS
Prototype of the
new version
of ETGs

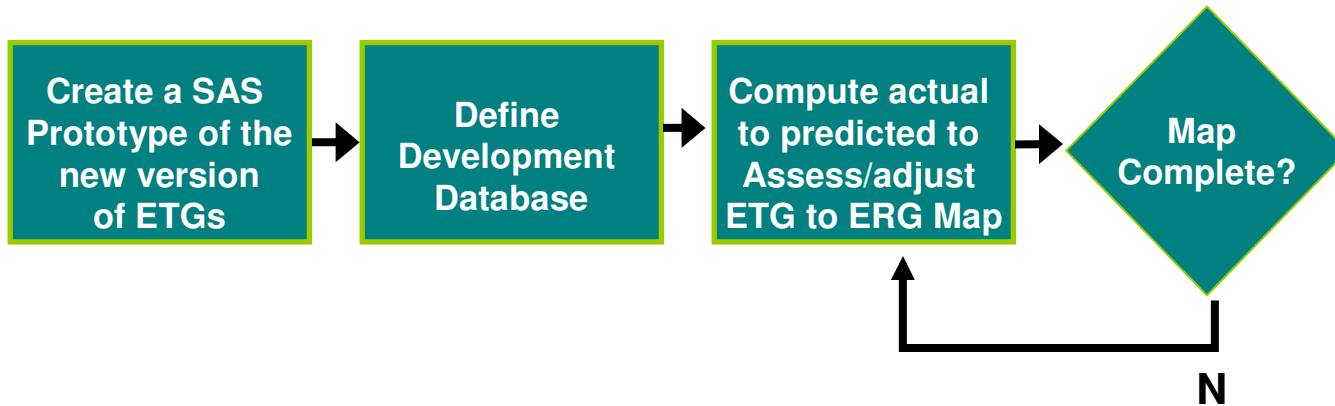
Prototype includes the new features planned for the upcoming release

Development Methodology – Step 2



- Geographical representation
- Time frame: 30 months
- Standard pricing – removes regional variation
- Large Volume – 8M members
- Create a validation data set (2M members)

Development Methodology – Step 3



- Iterate through until satisfied with ETG to ERG map

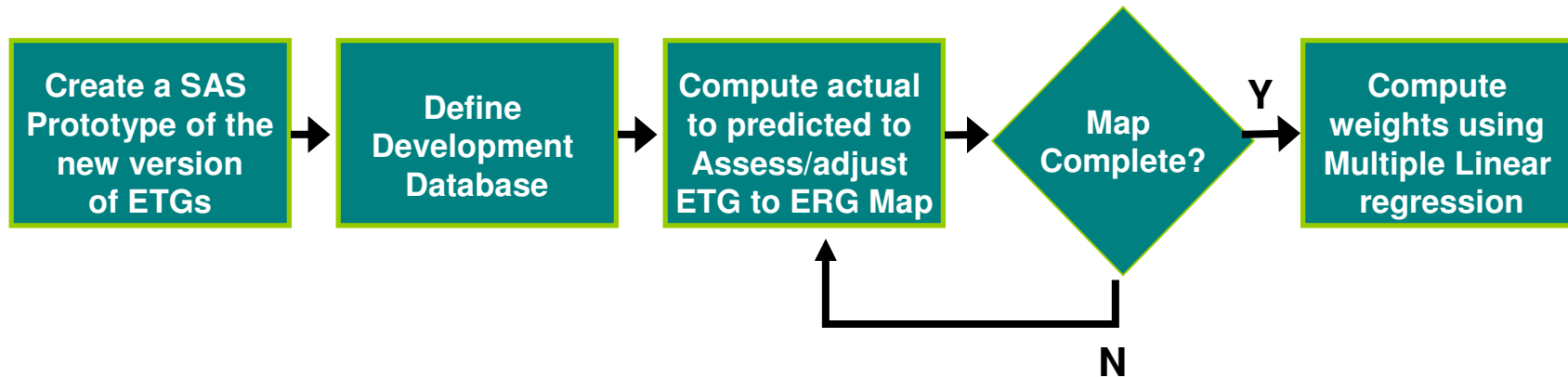
ETG to ERG Analysis

Base ETG	TI	Sev	ERG Desc	N	PMPM	Pred PMPM	Ratio
130600-Minor infect disease	0	1	Lower cost infectious disease	49,259	136	139	1.0243
130600-Minor infect disease	0	2	Lower cost infectious disease	72,715	299	297	0.9929
139900-Infect disease S&S	0	1	Lower cost infectious disease	185,968	242	241	0.9938
130600-Minor infect disease	0	3	Other moderate cost inf dis	30,908	569	569	1.0000
130400-Septicemia	0	1	Non HIV maj inf dis, I	3,410	1,412	1,528	1.0825
130800-Immunodeficiencies	0	1	Non HIV maj inf dis, I	2,698	1,162	1,145	0.9857
130800-Immunodeficiencies	0	2	Non HIV maj inf dis, I	484	1,365	1,043	0.7640
901200-Rx only-leprosy Tx	0	0	Non HIV maj inf dis, I	899	1,623	1,395	0.8595
130600-Minor infect disease	0	4	Non HIV maj inf dis, II	10,001	1,597	1,597	1.0000
130800-Immunodeficiencies	0	3	Non HIV maj inf dis, III	1,557	2,378	2,378	1.0000
130400-Septicemia	0	2	Non HIV maj inf dis, w signif cc/cb	5,718	2,746	2,736	0.9966
130400-Septicemia	0	3	Non HIV maj inf dis, w signif cc/cb	230	3,088	3,314	1.0731
130100-AIDS	0	1	AIDS/HIV	6,693	1,275	1,212	0.9506
130200-HIV sero positive wo AIDS	0	1	AIDS/HIV	877	1,058	1,254	1.1849
901100-Rx only-HIV/AIDS antiviral Tx	0	0	AIDS/HIV	1,325	1,060	1,250	1.1789
130100-AIDS	0	2	AIDS/HIV w signif cc/cb	3,561	1,894	1,892	0.9986
130100-AIDS	0	3	AIDS/HIV w signif cc/cb	539	3,269	3,288	1.0057

ETG→ERG Map: Infectious Disease MPC

Base ETG	TI	Sev	ERG	MPC	Hier	Priority	ERG Desc
130600-Minor infect disease	0	1	01.011	01	01	07	Lower cost infectious disease
130600-Minor infect disease	0	2	01.011	01	01	07	Lower cost infectious disease
139900-Infect disease S&S	0	1	01.011	01	all		Lower cost infectious disease
130600-Minor infect disease	0	3	01.021	01	01	06	Other moderate cost inf dis
130400-Septicemia	0	1	01.031	01	ign		Non HIV maj inf dis, I
130800-Immunodeficiencies	0	1	01.031	01	01	05	Non HIV maj inf dis, I
130800-Immunodeficiencies	0	2	01.031	01	01	05	Non HIV maj inf dis, I
901200-Rx only-leprosy Tx	0	0	01.031	01	01	05	Non HIV maj inf dis, I
130600-Minor infect disease	0	4	01.032	01	01	04	Non HIV maj inf dis, II
130800-Immunodeficiencies	0	3	01.033	01	01	03	Non HIV maj inf dis, III
130400-Septicemia	0	2	01.034	01	ign		Non HIV maj inf dis, w signif cc/cb
130400-Septicemia	0	3	01.034	01	ign		Non HIV maj inf dis, w signif cc/cb
130100-AIDS	0	1	01.041	01	01	02	AIDS/HIV
130200-HIV sero positive wo AIDS	0	1	01.041	01	01	02	AIDS/HIV
901100-Rx only-HIV/AIDS antiviral Tx	0	0	01.041	01	01	02	AIDS/HIV
130100-AIDS	0	2	01.042	01	01	01	AIDS/HIV w signif cc/cb
130100-AIDS	0	3	01.042	01	01	01	AIDS/HIV w signif cc/cb

Development Methodology – Step 4

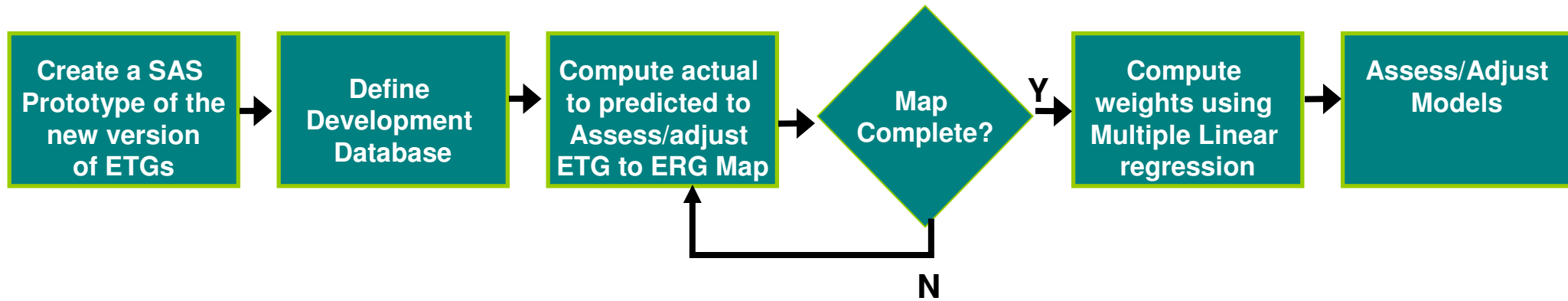


- Multiple linear regression is used to compute ERG weights
 - Model: $\text{Cost} = F(\text{age, gender, ERG1, ERG2, \dots, ERG}_{kkk})$
 - Separate models for each feature:

Input / Output
Partial Enrollment
Retrospective
Demographic

Truncation
A/U
Prospective

Development Methodology – Step 5



- All weights are examined for reasonableness
- Comparison within and between models – for consistency

Development Methodology – Step 5 cont.

- Model Metrics
 - R^2
 - Percent of variation that the model explains (larger is better)
 - Ranges between 0% and 100%
 - RMSE (Root Mean Square Error)
 - Measure of variation - standard deviation (smaller is better)
 - MAPE (Mean Absolute Prediction Error)
 - Compute absolute value of actual costs to predicted costs and average across the population (smaller is better)
 - Epidemiologic characteristics of identifying “large” claimants
 - Sensitivity, Specificity, PPV, NPV

How did we do?

The performance of our predictive models can be measured by R-square – the percent of variation the model accounts for. This is done by correlating what we predicted to happen against what actually happens. This is just one example of many metrics than be used to assess our work (MAPE, Predictive Ratios, etc.)

Model/Population	6.0	7.0	Absolute Increase	% Increase
100K Medrx to Medrx-Prospective	23.10%	24.40%	1.30%	5.60%
100K Medrx to Medrx-Retrospective	51.20%	56.50%	5.30%	10.40%
25K Medrx to Medrx-Prospective	29.70%	29.90%	0.20%	0.70%
25K Medrx to Medrx-Retrospective	54.10%	57.40%	3.30%	6.10%
A/U 100K Medrx to Medrx-Prospective	19.00%	20.00%	1.0%	5.30%



Q & A