The First Annual Meeting of

The PMRG Institute

The Art and Science of Healthcare Marketing Research

September 16-18, 2007
Philadelphia, PA

Hidden Gold: Finding new insights in data you already own

Graham Crawford
Senior Director

Pieter Sheth-Voss, PhD
Research Director

EIDETICS

Market research can often yield insights even after the "final" presentation

- Brand teams spend hundreds of thousands of dollars on market research
- Data must often be applied to questions beyond those that initially prompted the research
  - New questions arise after study completed
  - Need data cut differently than final report
  - Need to drill into sub-populations where data get thin
  - Need to link data on different aspects of the market
- Most questions start with "why does…?" and most answers start with "it depends. Here's the story…"

Brand teams often have many sources of data

MD survey
- Clinigraphics
- Treatment attitudes
- Product perceptions

Conjoint
- Product profile
- Product interest
- Self-explicated ratings

Patient charts
- Demographics
- Treatment history
- Comorbidities
- Disease history
- Current regimen
- Outcome

Sales force
- Details
- Product interest

Rx data
- Decile
- Rx segment

Longitudinal
- Adherence
- Cost of care

Epidemiology
- Incidence, prevalence
- Trends

Go beyond crosstabs and final reports to make better use of existing data

- Use interactive analytics to answer new queries
- Use models to drill deep even where data are thin
  - Modeling allows researchers to leverage the full data set
  - Bayesian networks are a natural extension of conjoint
  - Derive segmentations to identify customer differences
- Link datasets together based on common constructs
  - "Hard" versus "soft" links
  - Bayesian classifiers allow projection to other datasets

Employ interactive analytics with elemental data

- Being able to easily cut elemental data in new ways increases the potential value of your data
- Interactive analytic software allows you to efficiently cut the data in new ways to support new queries
  - MarketSight®
  - Tableau®
  - Spotfire®
  - Provenance™
- Request access to the data in elemental form
  - Addition (not substitute) for final reports

Case study:
CDC National Ambulatory Medical Care Survey (NAMCS)

Overview of the CDC NAMCS database

- Annual survey of national outpatient medical care
  - Annually since 1973 except 1982–1988
  - One of series of related studies
- Broad selection criteria
  - Office-based physicians in direct patient care
  - 70+% invite-to-participation rate (without honorarium)
- Physicians fill out case forms for patient visits
- Public domain, basis for many published articles
  - 25,000 patient cases
  - 200 "primary" variables + additional derivatives

Overview of NAMCS data

Physician info

- Specialty
- Clinigraphics
- Difficulty referring
- Consults
  Visits
- Sources of revenue

Patient info

- Patient information
- Reasons for visit
- Continuity of care
- Counseling
  Providers seen
  Disposition
- Diagnoses
  Causes of injury
- Medications /
  Surgery

Data as distributions

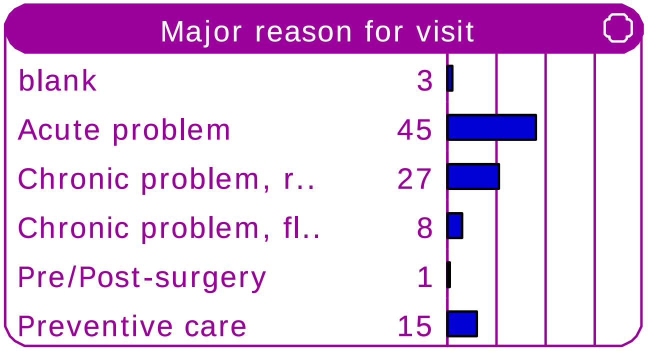
In market research, the individual answers from each respondent are less important than the generalizations we can make about the distribution of responses

Each **question** from the research..

What is the major reason for this visit?

- ❍ acute problem
- ❍ chronic problem, remission
- ❍ chronic problem, flare
- ❍ pre/post surgery
- ❍ preventive care

… is best viewed as a **variable** with a **distribution** of responses



There is often no one “right” way to crosstab a field for all future queries. For example, “date of visit"…

Visit date

Year of visit	
2003	1
2004	99

By year

Month of visit	
january	8.8
february	8.9
march	9.2
april	8.4
may	9.9
june	8.7
july	7.8
august	6.7
september	7.9
october	8.2
november	8.2
december	7.5

By month

Day of week of visit - recoded	
sunday	0
monday	23
tuesday	23
wednesday	19
thursday	20
friday	13
saturday	1

By day of week

And it is easy to find far more complex variables … The NAMCS data has rich detail on the Rx decision

Open-end responses (coded)

Over 4000 different medications

And up to 8 meds may be listed for each patient!

Diagnoses

- Patient's reason for visit (broad)
- Patient's reason for visit
- Diagnosis (detailed)
- Diagnosis (broad)
- Cause of injury (broad)
- cause of injury (detailed)

Decisions

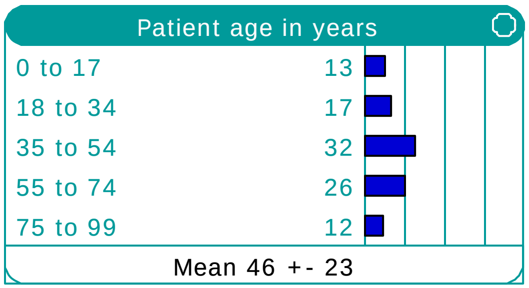
- Medication(s)
- Generic code(s)
- NDC code(s)
- Ingredient code(s)
- Surgical procedure - 2 digit
- Surgical procedure - 3 digit
- Surgical procedure - 4 digit

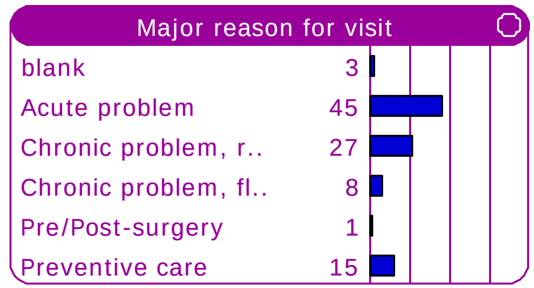
Multiple levels of detail

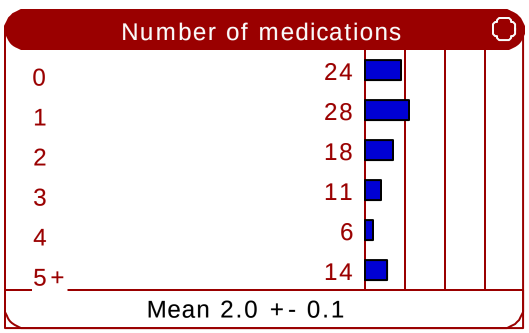
Drill into subgroups for key questions

- e.g., "Top differences between FP/GPs and IMs?"

FP/GP only







Top 5 agents prescribed

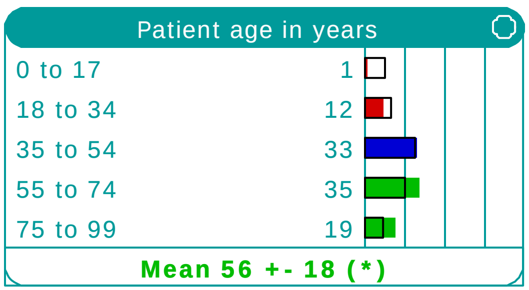
FP / GP	IM
LIPITOR	LIPITOR
LISINOPRIL	A.S.A.
ALBUTEROL	SYNTHROID
ATENOLOL	ATENOLOL
SYNTHROID	PREVACID

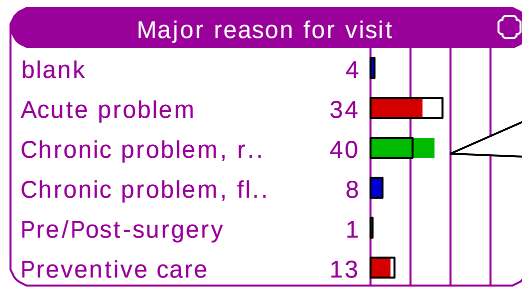
Drill into subgroups for key questions

- e.g., "Top differences between FP/GPs and IMs?"

IM compared to FP/GP only

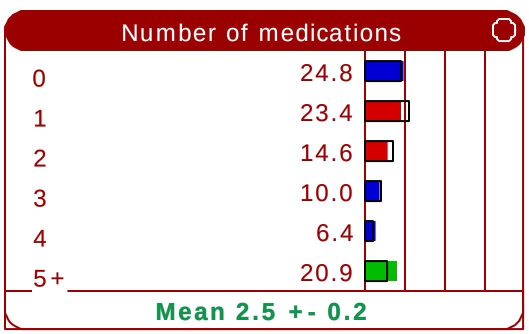
IMs' patients are older...





IMs treat more chronic conditions

IMs prescribe more meds



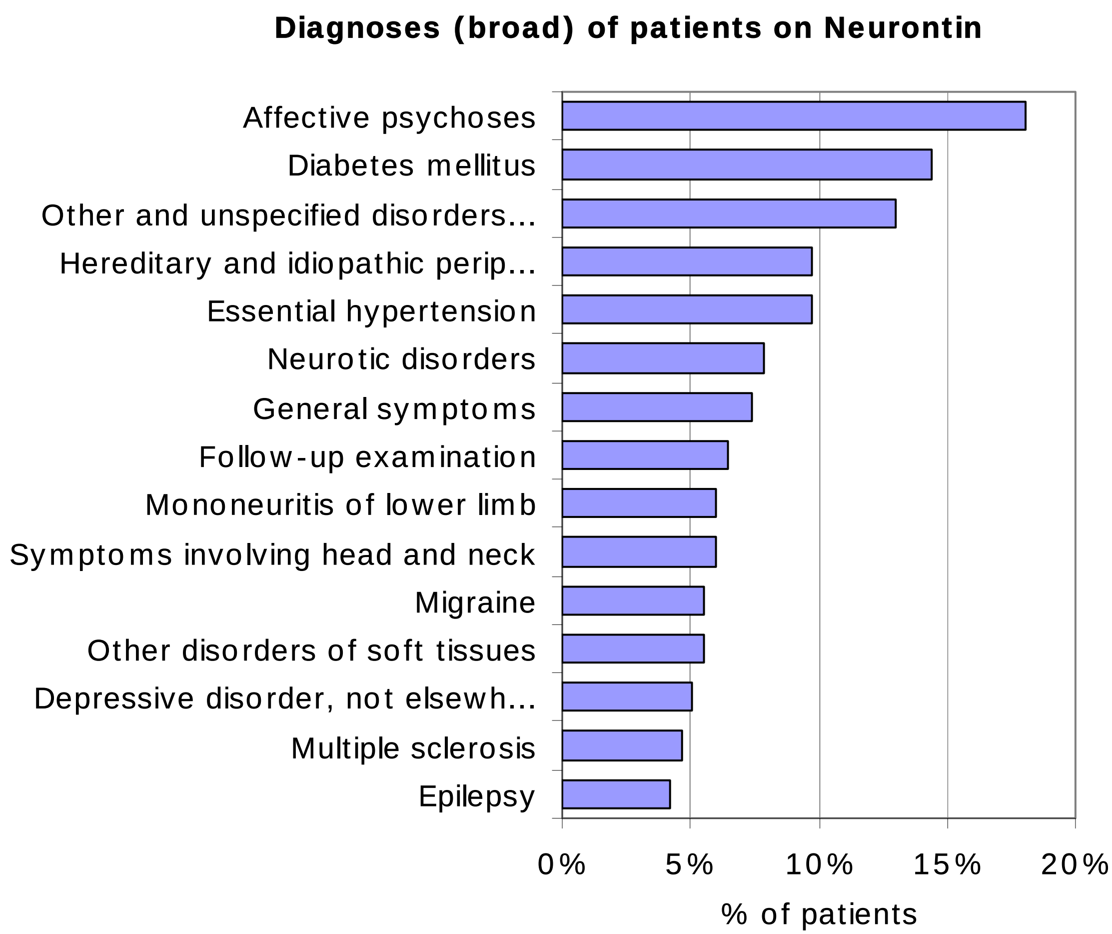
Top 5 agents prescribed

FP / GP	IM
LIPITOR	LIPITOR
LISINOPRIL	A.S.A.
ALBUTEROL	SYNTHROID
ATENOLOL	ATENOLOL
SYNTHROID	PREVACID

e.g., “What diagnoses do Neurontin patients have?”

D	Physician specialty recode
General/family prac..	10.6
Internal medicine	6.8
Pediatrics	0.0
General surgery	1.7
Obstetrics and gyn..	0.3
Orthopedic surgery	0.3
Cardiovascular dis..	4.8
Dermatology	0.3
Urology	0.7
Psychiatry	18.5
Neurology	44.5
Ophthalmology	0.3
Otolaryngology	1.4
Other specialties	9.6

Patient sex	
female	59
male	41

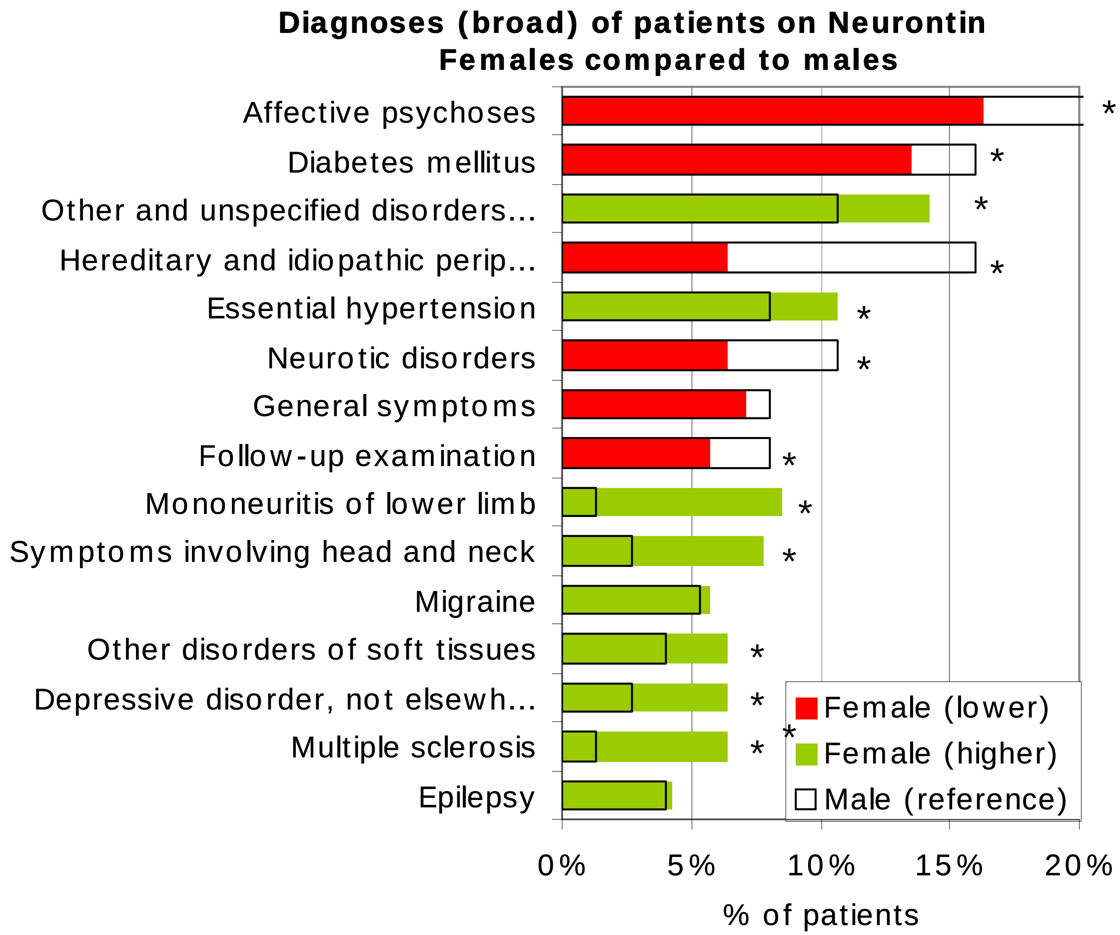


PMRG

e.g., “How do diagnoses differ between female vs. male Neurontin patients?”

D Physician specialty recode	
General/family prac..	12.0
Internal medicine	6.2
Pediatrics	0.0
General surgery	2.6
Obstetrics and gyn..	0.5
Orthopedic surgery	0.5
Cardiovascular dis..	4.2
Dermatology	0.5
Urology	0.0
Psychiatry	15.6
Neurology	46.9
Ophthalmology	0.5
Otolaryngology	1.0
Other specialties	9.4

Patient sex	
female	100
male	0



PMRG

Case study:
Primary field research

Patient chart audit and MD-level survey

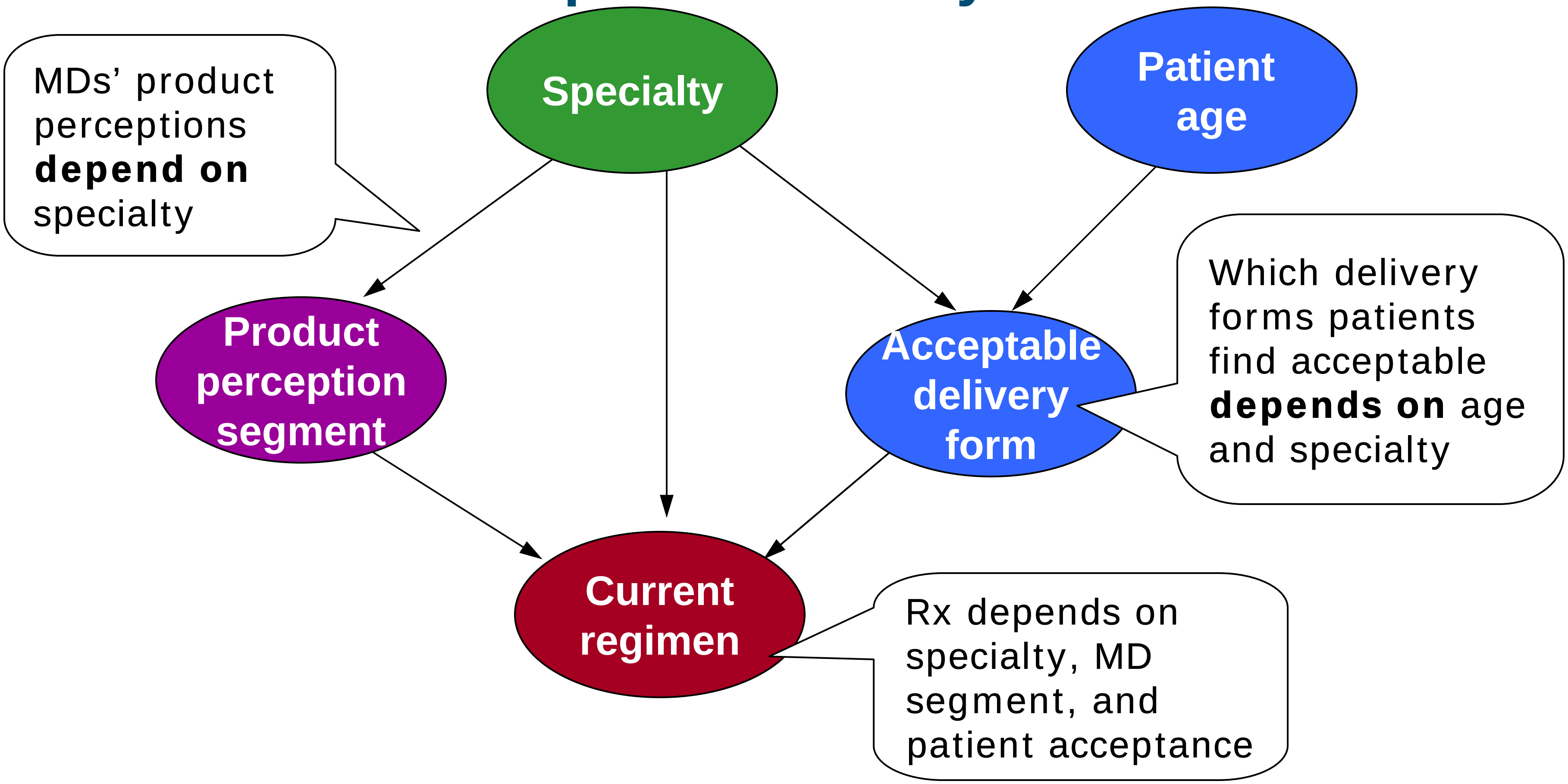
- Client is an established leader seeking to compete with recent and new product entries
- Therapeutic area disguised as “male pattern baldness” to convey ideas yet maintain confidentiality
- Brand team wanted a detailed understanding of how patient, physician, product, and setting factors drive treatment decisions in this therapeutic area
- Methodology:
  - Detailed patient chart review
  - MD survey on clinigraphics, attitudes, and product ratings

"Each patient is unique, like a snowflake" – MD

Bayesian networks uncover *how* "it depends"

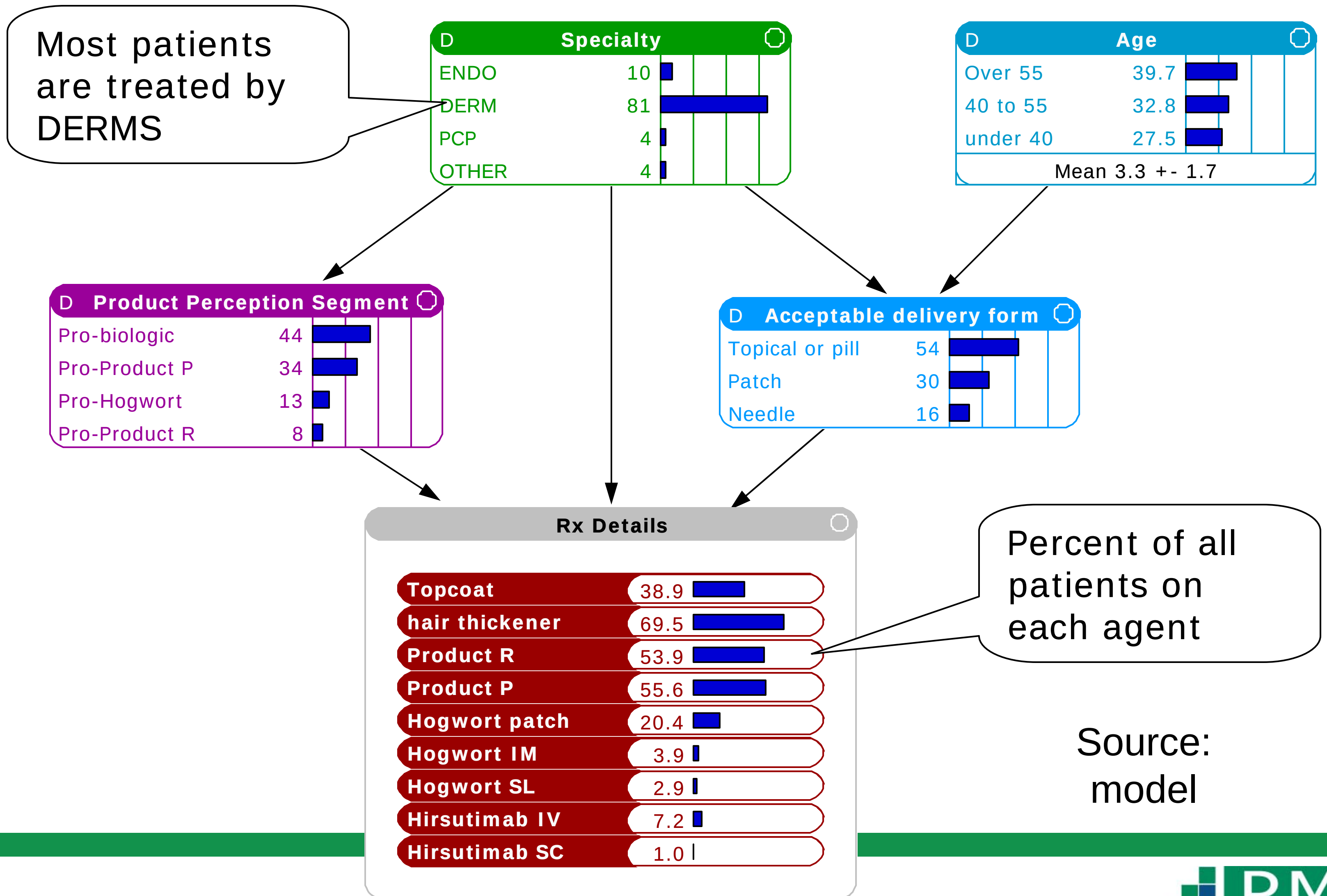
- Probabilistic foundation for artificial intelligence
- Many applications, e.g., medical expert systems
- "Bayesian networks" coined by Pearl in the 1980s
- Extensively researched at Stanford, Microsoft, etc.
- Available commercially in many packages today
- Superset of many classic market research models:
  - Decision trees
  - Conjoint models
  - Latent class segmentation

Statistical analysis can identify inter-relationships, which we can represent as Bayesian networks



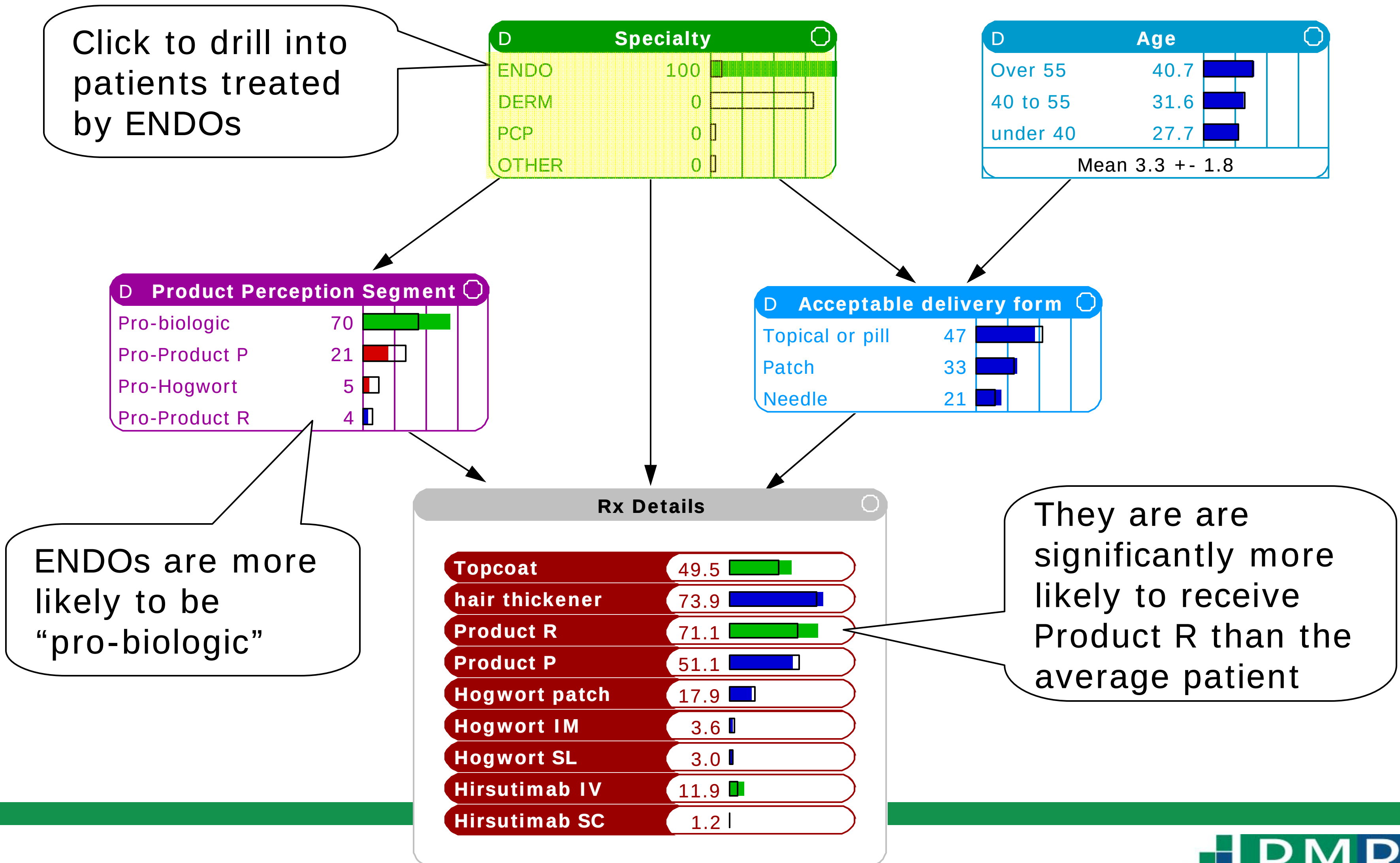
Beneath each node is a conjoint model of how it depends on each driver

Bayesian networks also simulate the models…

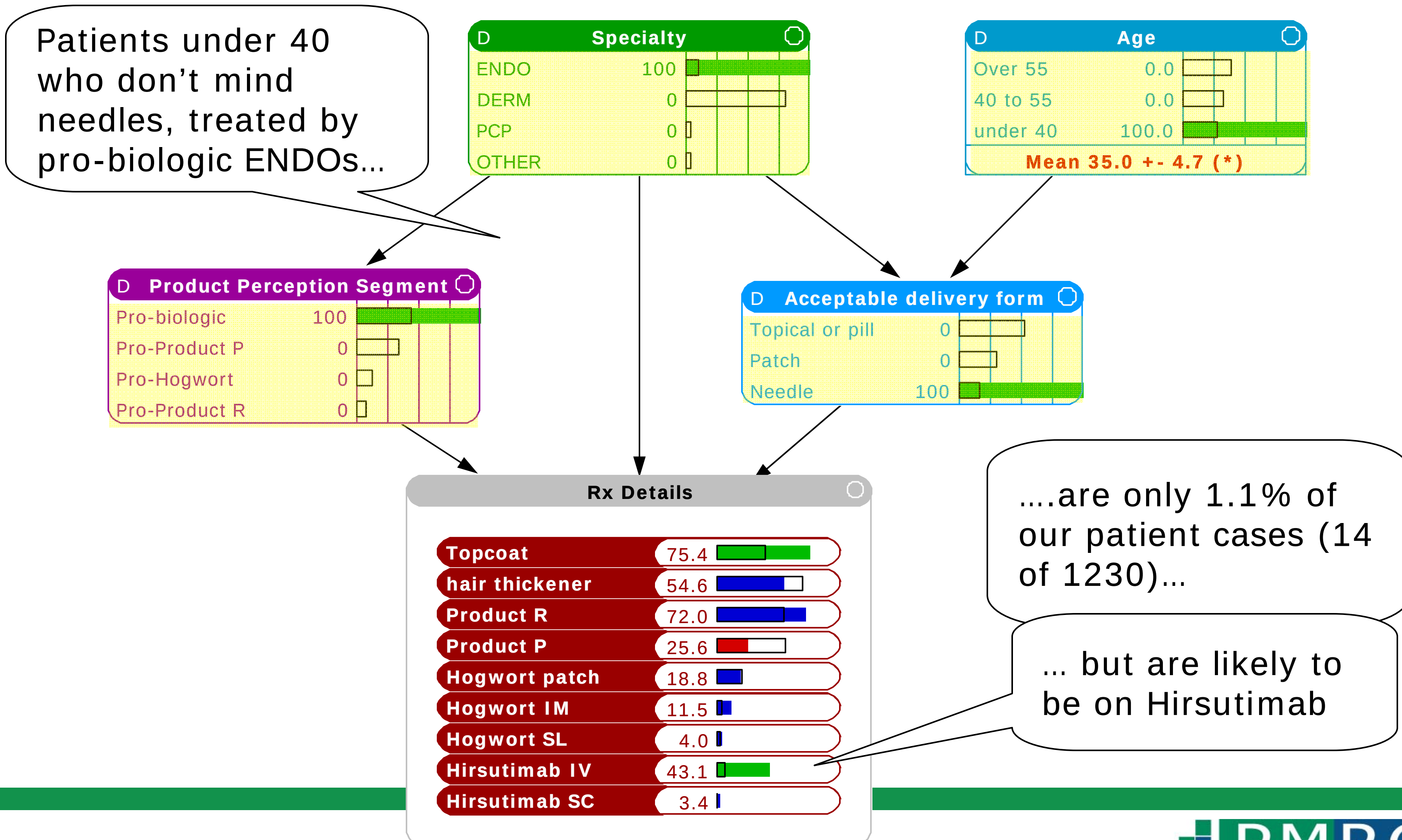


Source: model

…and let you interactively query them

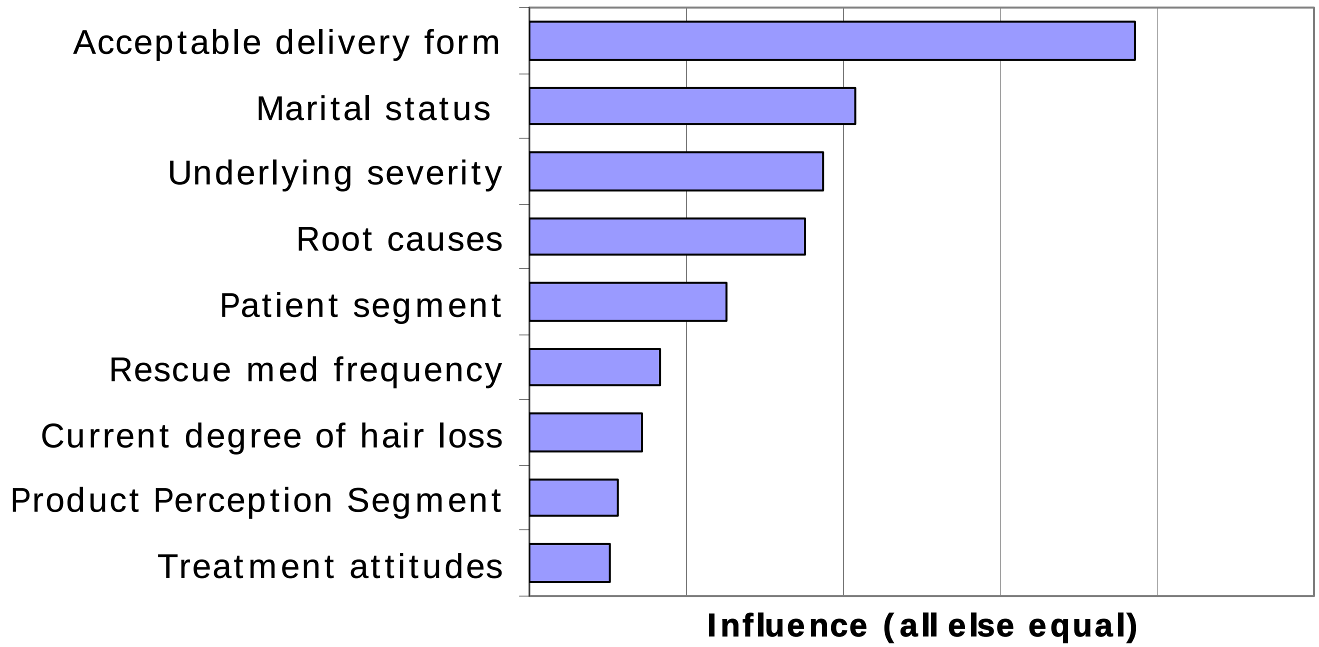


You can use a model to drill down quite deep…

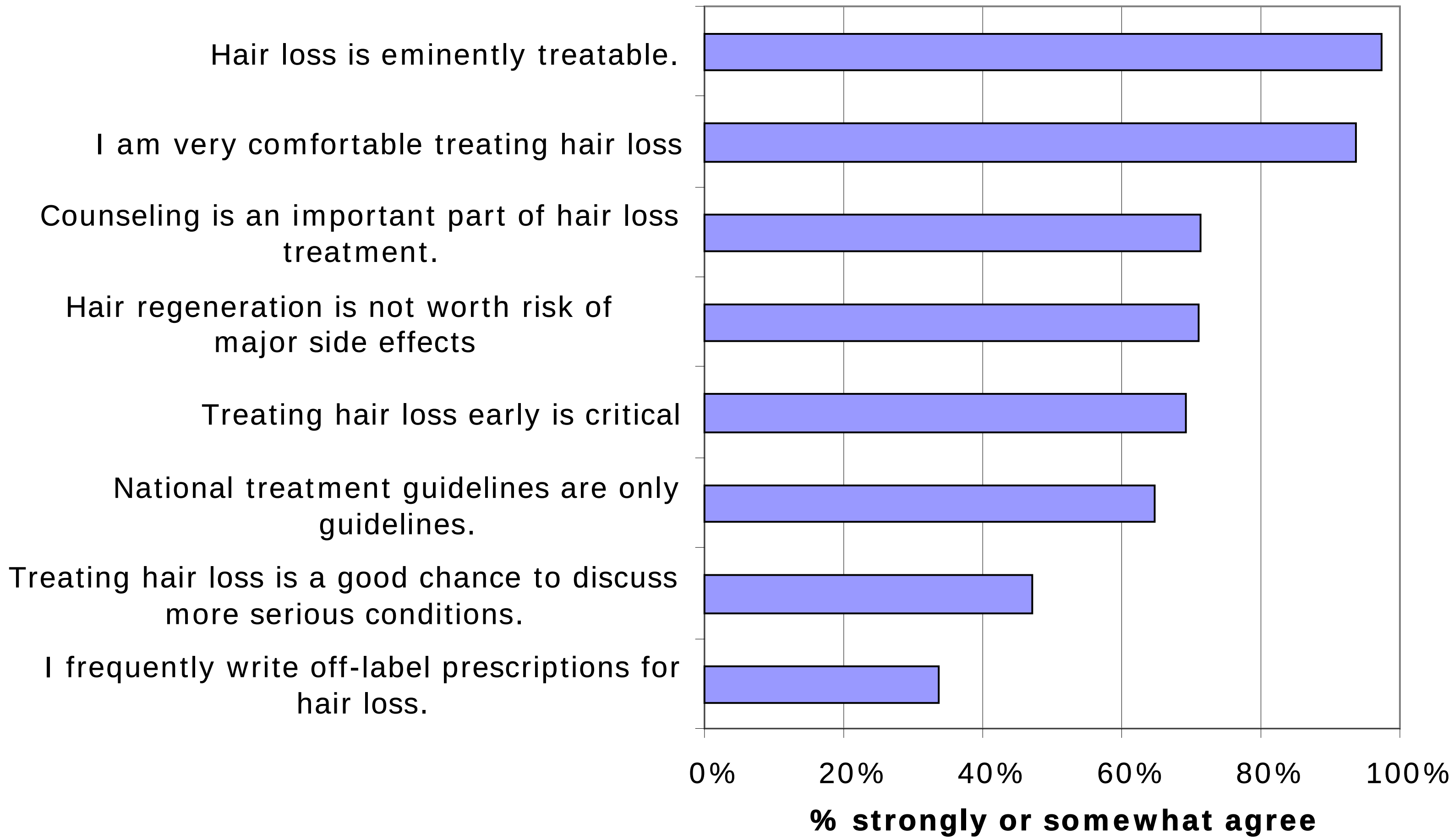


Just as with conjoint models, you can quantify how each factor depends on its drivers

Importance of key factors driving choice of hair loss regimen



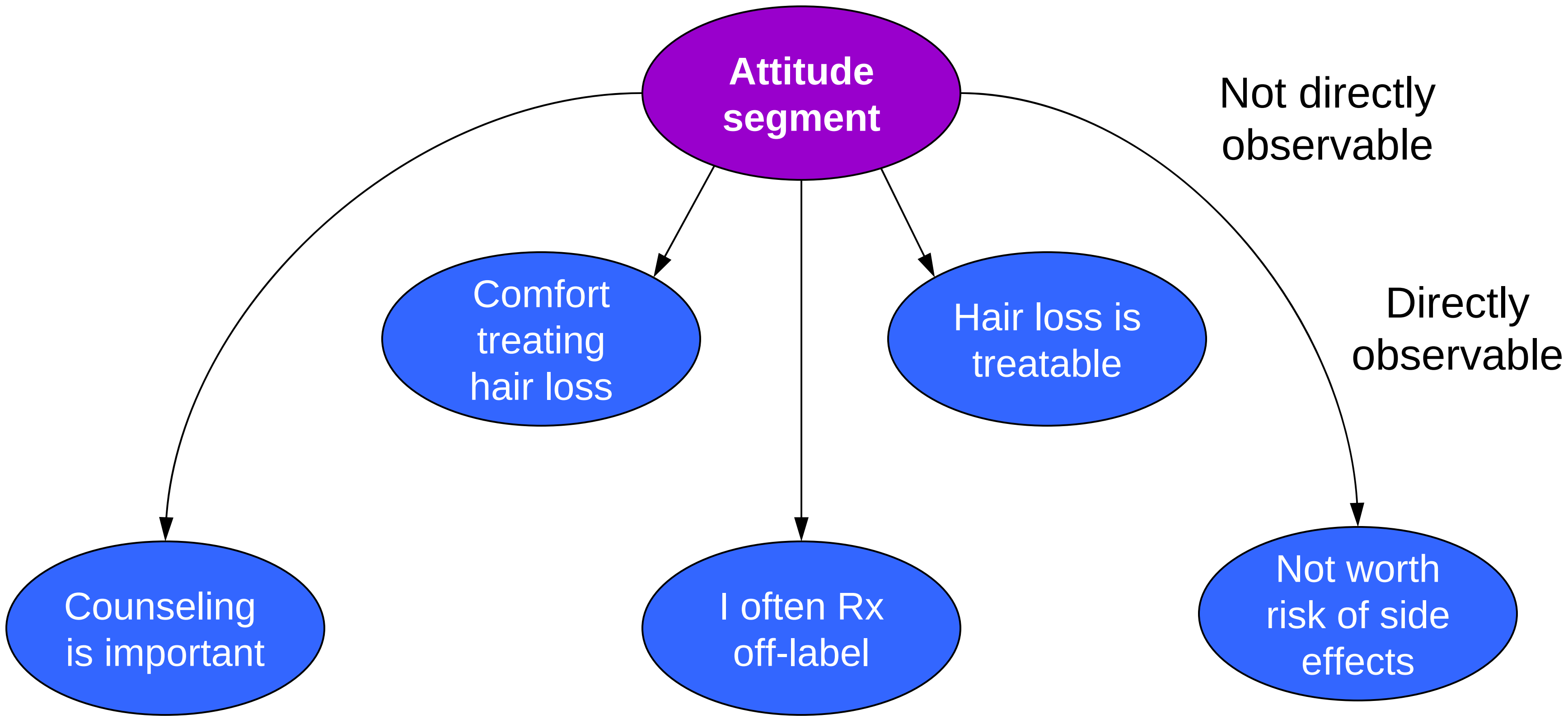
Surveys often contain collections of related measures, such as attitudinals



Beneath those top-level summaries may lie key differences

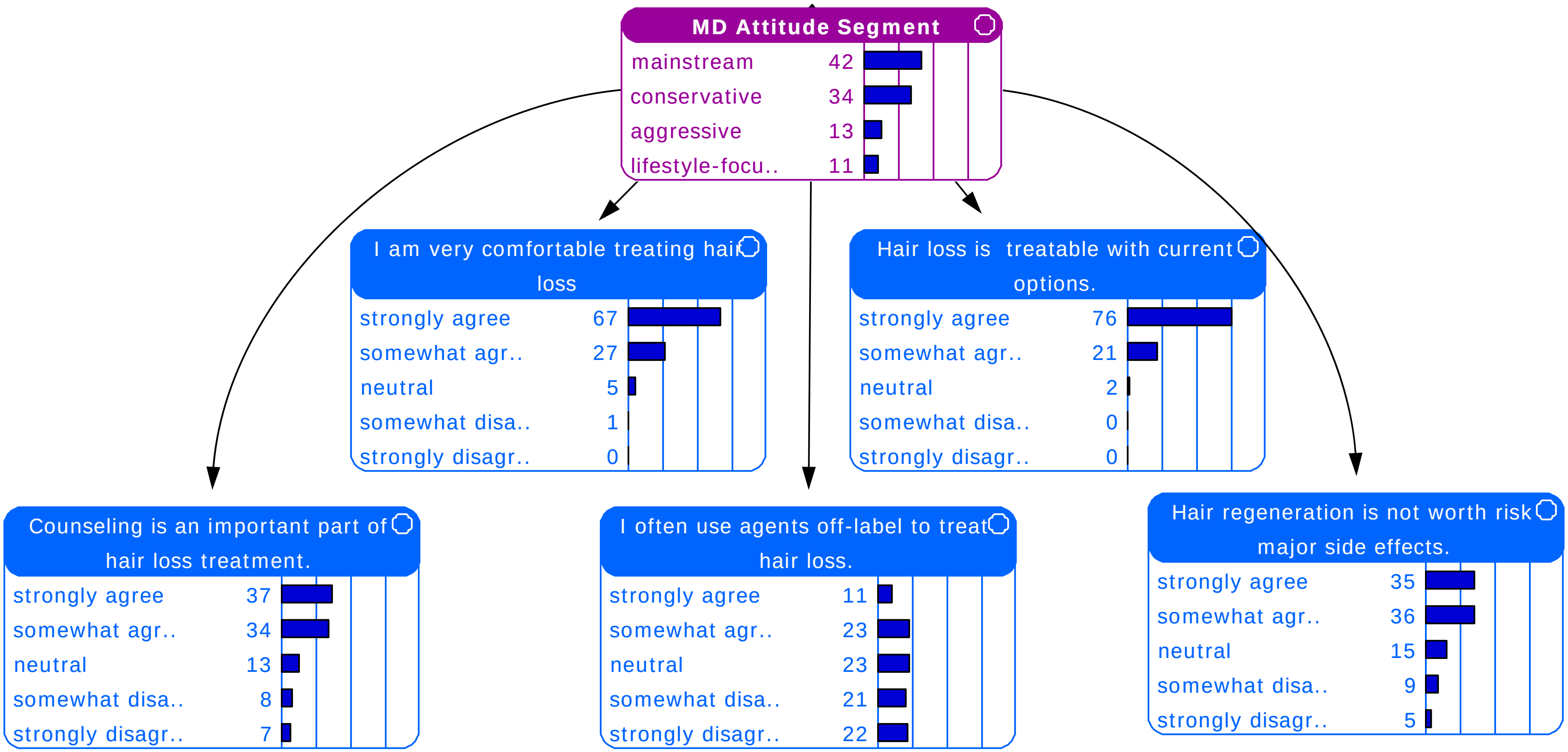
- Even if most respondents agree or disagree with one particular attitude, not *everyone* does
- Items within a battery frequently have complex correlations with each other and other MD attributes
- Practically, a segmentation is a variable that explains the correlation among important variables
- Segmentation analysis can often distill items into new constructs that suggest key insights
- Such segments can “capture” much of the information in the items…and be projected to related datasets

We often infer that observable data are correlated because they all depend on a hidden "segment" variable

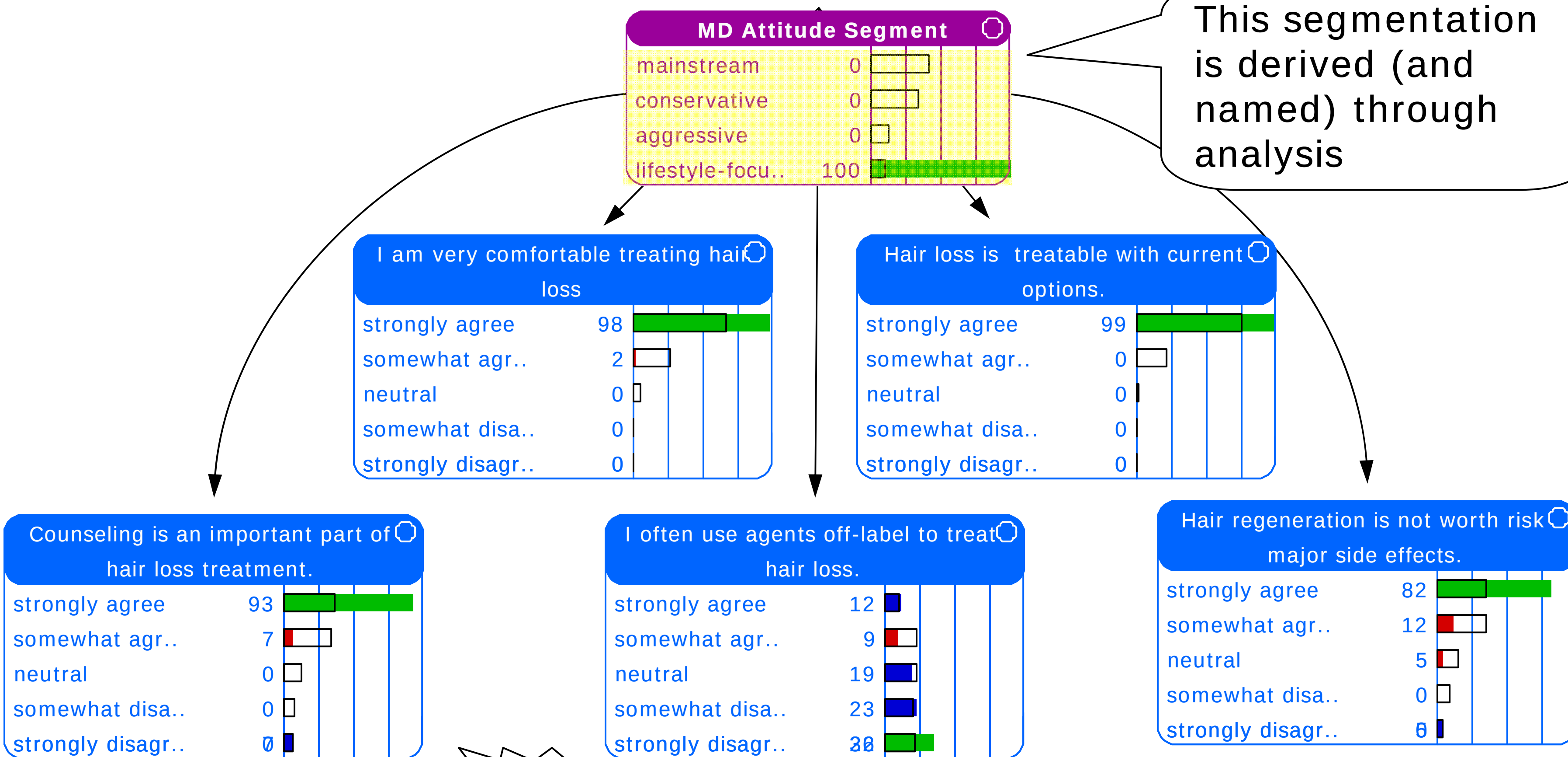


The attitude segmentation can now be used as a shorthand for attitudes in other analyses

Bayesian networks represent segmentations well (and are exact for latent class segmentations)



The segment "captures" information about the items: i.e., knowing segment, you can infer the attitudes

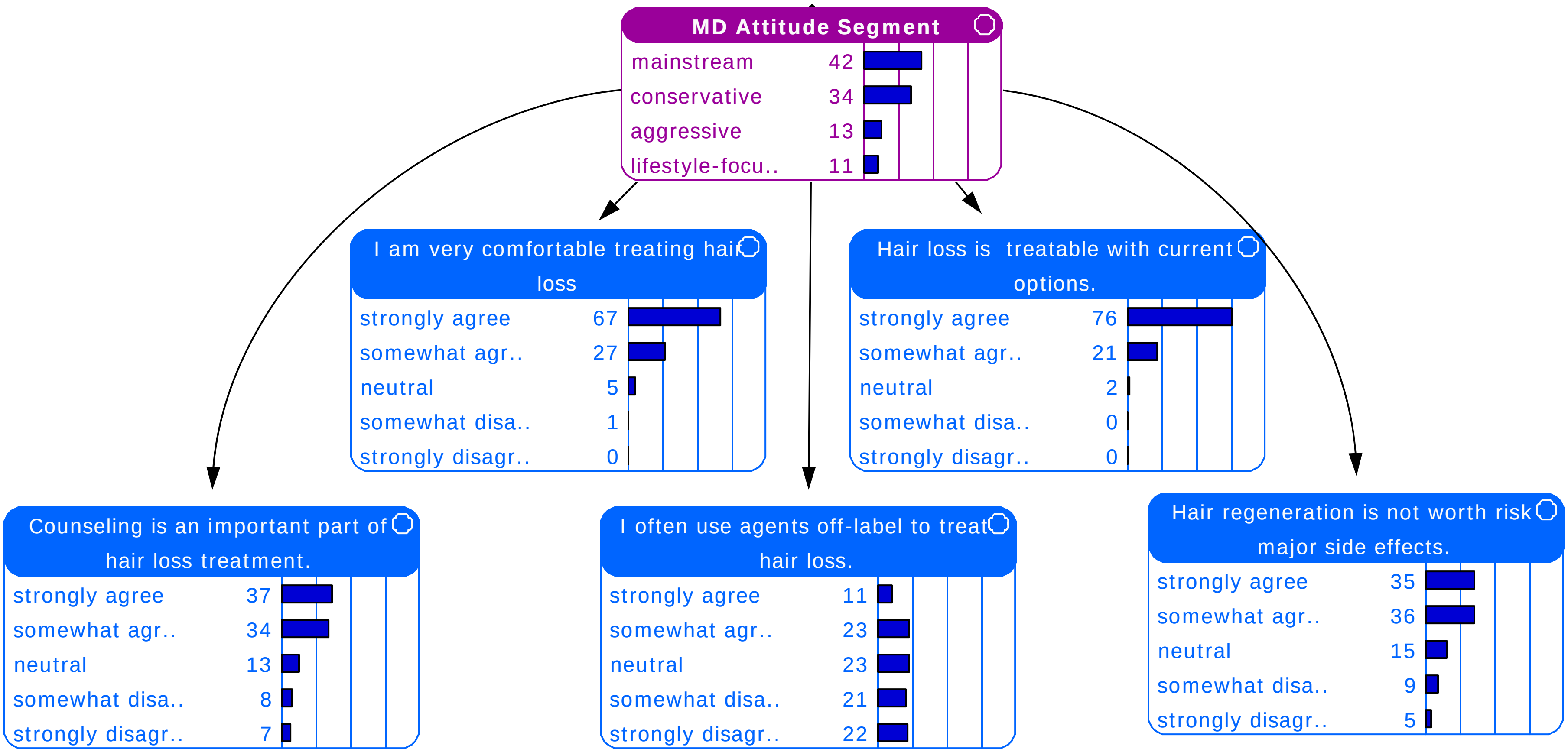


This segmentation is derived (and named) through analysis

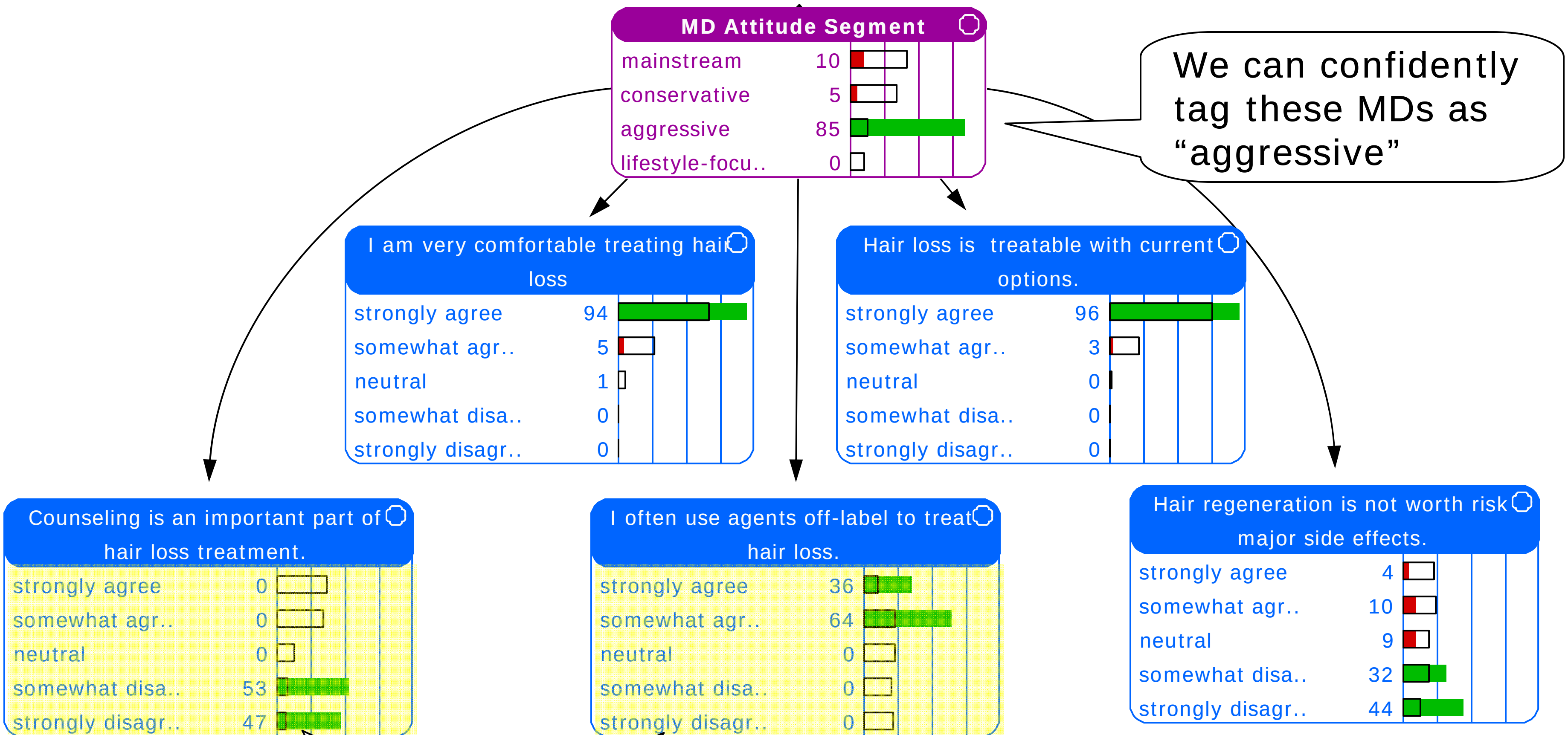
These variables are used to derive the segmentation

PMRG

And *vice versa*: knowing one or more items, you can infer the segment, via Bayesian classification



And *vice versa*: knowing one or more items, you can infer the segment, via Bayesian classification



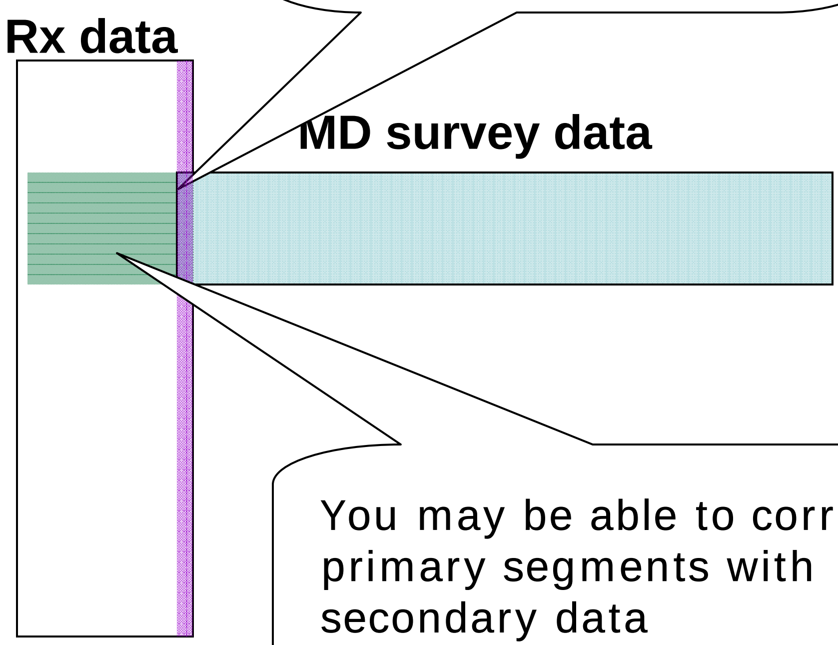
If just these two values are observed….

Note: Bayesian classifiers can classify physicians using variables beyond those used to derive the segmentation

Related studies can often be linked to leverage insights if they have respondents and unique IDs in common

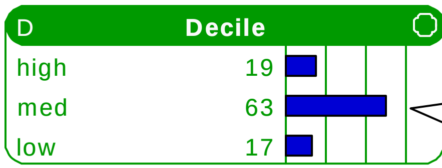
If both datasets overlap in both respondents **and** identifying data (e.g., ME Number…)

"Hard" links



You may be able to correlate primary segments with secondary data

For example, in our case study, attitudinal segment is somewhat related to decile (controlling for specialty)



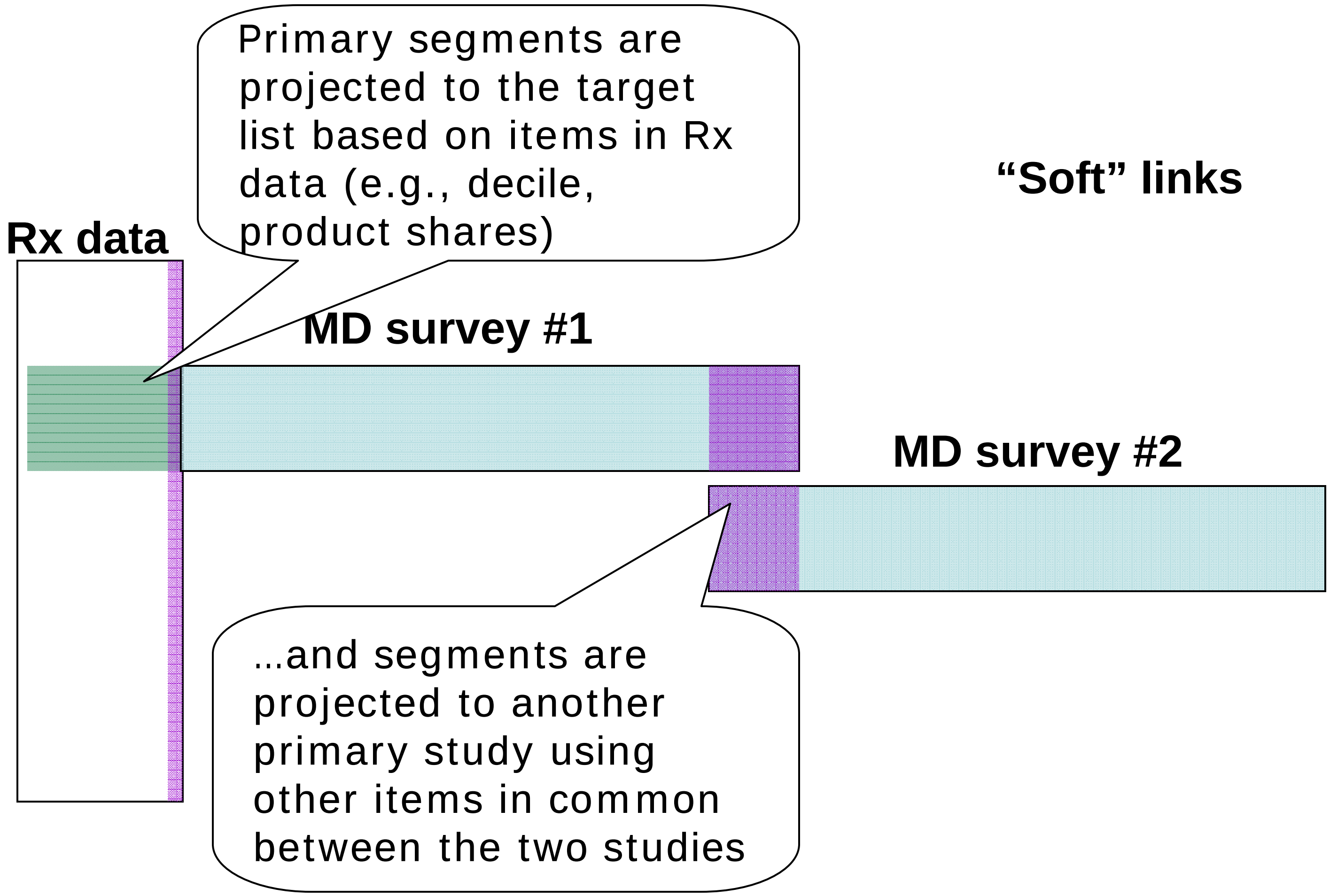
We linked the primary research to secondary data using IMS Prescriber ID

	MD Attitude Segment			
Decile	mainstream	conservative	aggressive	lifestyle-focus
high	21%	11%	16%	48%
med	63%	69%	62%	48%
low	16%	20%	22%	4%

Lifestyle-focus MDs are clearly higher decile; conservative MDs are lower

PMRG

Classification tools may also project segmentations between studies if they share key items in common



Summary

- Market research data are too rich to expect the final report to answer every future query — use interactive analytics to efficiently access data in elemental form
- Use conjoint modeling to quantify relationships among variables, and Bayesian networks to synthesize the many individual models
- Segmentation can concisely capture information in complex surveys, and Bayesian classifiers can often project segments to other datasets
- Link datasets to leverage insights across studies

Important Caveats

- Interactive analytic software is increasingly intuitive and easy to use, but can require a learning curve
- Causal modeling requires a fair amount of well-formed data, not all datasets are suitable
- Linking datasets requires some overlap, either ID codes or content; linking is not always possible
- Differences among customers can be too complex to be conveyed well in a simple segmentation

For more information, please contact . . .

EIDETICS

Graham Crawford
Senior Director
617.254.4100
gcrawford@eidetics.com

Pieter Sheth-Voss
Research Director
617.254.4100
pvoss@eidetics.com