Medication Adherence: Tailoring the Analysis to the Data

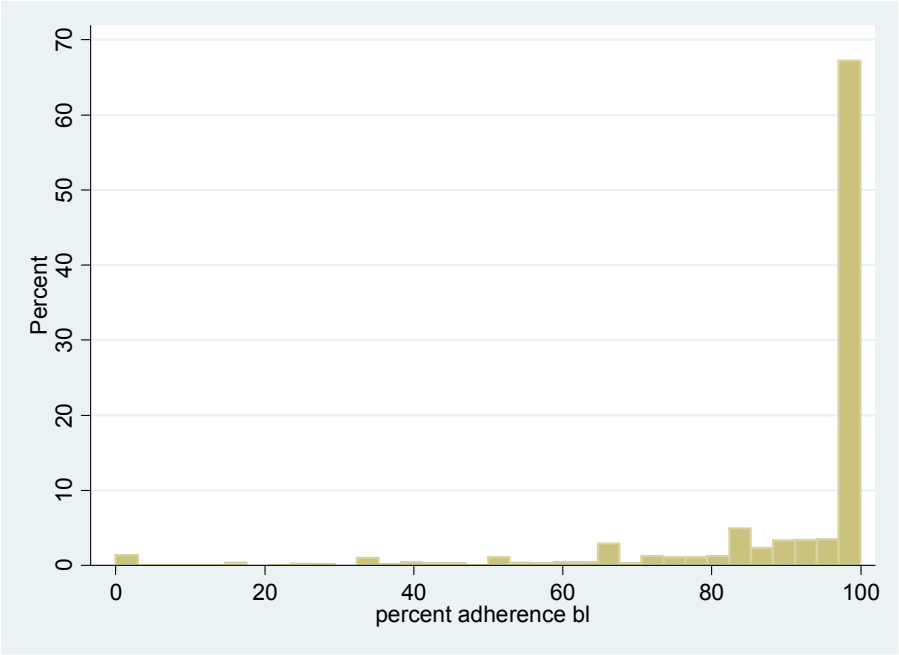
Parya Saberi, Pharm.D.

June 3, 2010

University of California
San Francisco
UCSF
AIDS Research Institute

Problems with Adherence

- Assessment → focus of many studies
- Analysis → not discussed much
  - As a continuous variable, adherence data are highly left-skewed and have a pile of values at 100%



Current Method of Analyzing ARV Adherence

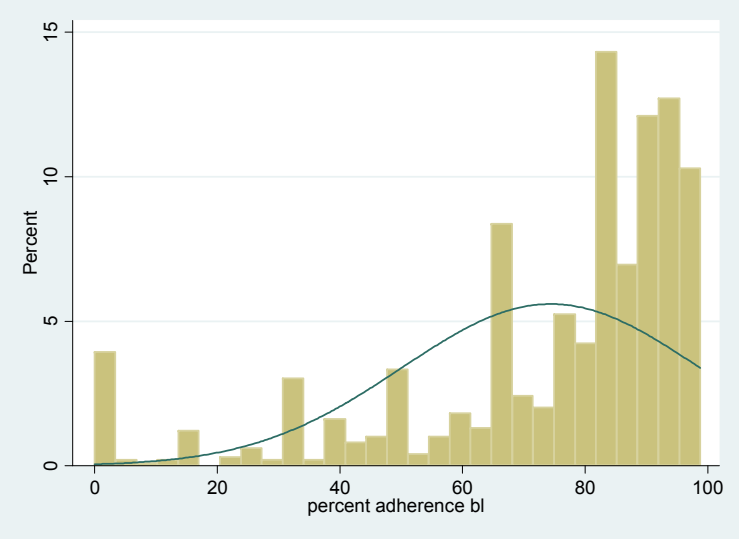
- Dichotomizing
  - May lead to loss of power
  - May lead to loss of valuable information
  - Usually determined post hoc
  - Requires choosing arbitrary & subjective cut-offs
    - 90%, 95%, or 100%
- Analyzing using logistic regression

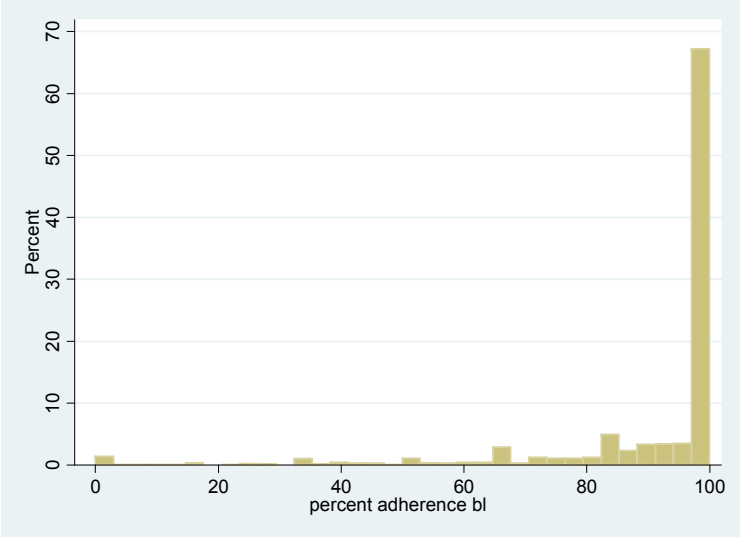
Objective

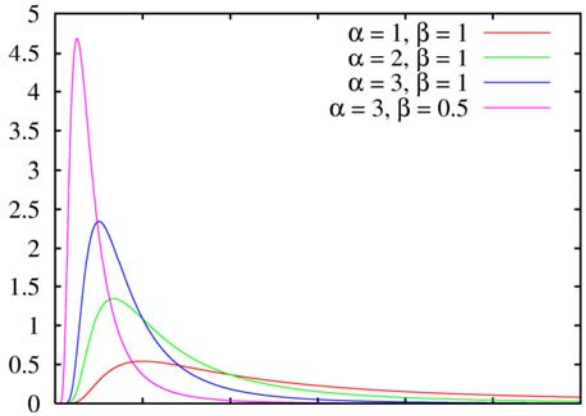
- Determine a complement to logistic regression to analyze medication adherence data, where adherence can be analyzed as a continuous variable.
- Demonstrate an example from actual ARV adherence data.
- Illustrate results of simulation models for a variety of scenarios.

Similarity Between the Shape of Adherence Data & Gamma Distribution

- Underlying reverse gamma distribution w/ inflation at 100%







http://upload.wikimedia.org/wikipedia/commons/1/1f/Inverse_gamma_pdf.png

Generalized Linear Models: Gamma Distribution

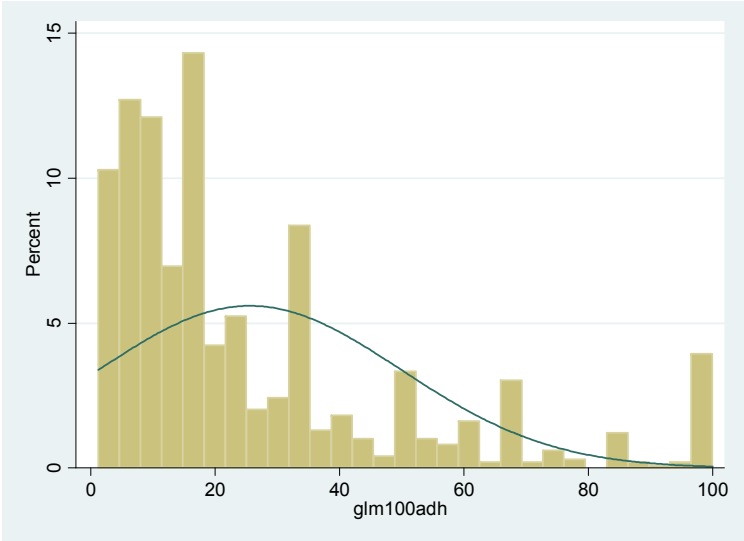
- Can be used when outcome is:
  - Continuous
  - Non-negative
  - Right-skewed
- Model assumption:
  - Standard deviation of outcome is proportional to mean

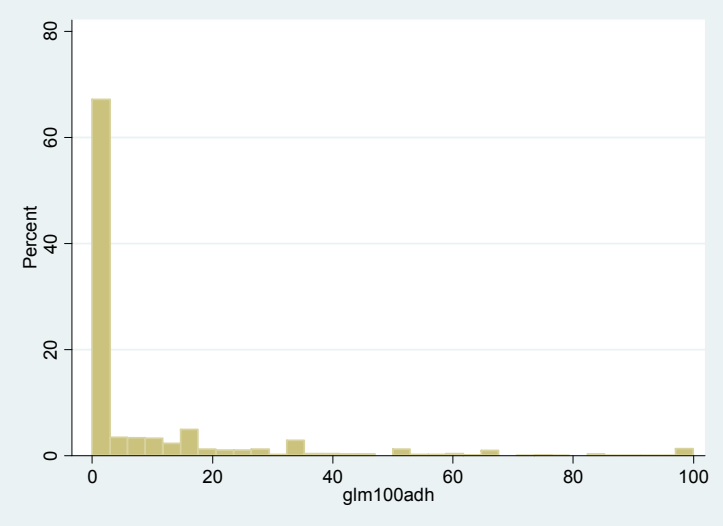
Applying Gamma Distribution to Adherence < 100%

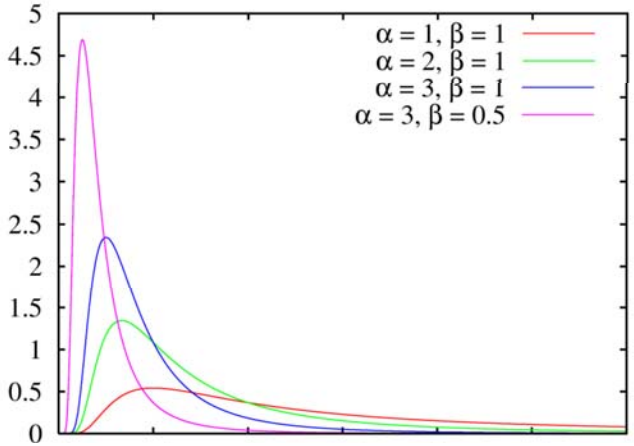
- Gamma distribution can be used when outcome is:
  - √ Continuous
  - √ Non-negative
  - → Right-skewed: but adherence data are left-skewed
    - Can transform by subtracting adherence from 100% (i.e. 100 – percent adherence): resulting in **percent *non-adherence***

Transforming Adherence Data

- Underlying gamma distribution with inflation at 0%







http://upload.wikimedia.org/wikipedia/commons/1/1f/Inverse_gamma_pdf.png

EXAMPLE

- Performed secondary data analysis of baseline data of the **Healthy Living Project** (HLP)
  - Data on 2845 individuals on ARVs at baseline
  - Variables:
    - ARV adherence: self-reported ACTG 3-day adherence
    - Age: ≤34, 35-44, ≥45

Step 1: Logistic Regression

- Can dichotomize *non-adherence* at a cutoff of 0%

Step 2: GLM Gamma Distribution

- For all values > 0%, we can treat percent *non-adherence* as a continuous variable

Interpretation

- Because of transformation of outcome, ratios and coefficients are in terms of increasing *lack of adherence* or *non-adherence*

EXAMPLE

- Association between age and ARV non-adherence

Logistic Regression

Age category	Frequency (%)	Percent with 0% Non-adherence	Odds Ratio (95% CI)	p-value
≤34	420 (14.8)	64.8	Ref	-
35-44	1462 (51.4)	62.4	1.11 (0.88, 1.39)	0.39
≥45	962 (33.8)	69.4	0.81 (0.63, 1.03)	0.09

Generalized Linear Model

Age category	Frequency (%)	Mean % Non-adherence	Coefficient (95% CI)	p-value
≤34	148 (14.9)	30.1	Ref	-
35-44	549 (55.4)	24.7	0.82 (0.70, 0.96)	0.02
≥45	294 (29.7)	25.0	0.83 (0.69, 0.99)	0.04

Interpretation of Logistic Regression

- Odds of *non-adherence* for individuals ≥ 45 years of age is 0.81 (95% CI = 0.63-1.03, p = 0.09) times higher than individuals ≤ 34 years of age.

Interpretation of GLM

- Among individuals with > 0% non-adherence:
  - those ≥ 45 years of age have a 0.83 (95% CI = 0.69-0.99, p = 0.04) times higher risk of non-adherence than those ≤ 34.
  - mean predicted *non-adherence* for individuals ≤ 34, 35-44, and ≥ 45 years of age is 30.1%, 24.7%, and 25.0%, respectively.

Simulation Model

- To determine type I error and power of use of GLM + logistic regression versus logistic regression alone, we simulated data and evaluated differences using:
  - **Model 1 (sample size model):** similar gammas and zero inflations as data in HLP but sample sizes of 200 to 2000
  - **Model 2 (gamma distribution model):** similar zero inflations as HLP, sample size of 2850, but changes in gamma distributions
  - **Model 3 (zero inflation model):** similar gamma distributions as HLP, sample size of 2850, but changes in zero inflations

Simulation Model

- Outcome: ARV non-adherence
- Predictor:
  - Hypothetical adherence predictor (categories 0-3)
  - Within boundaries of other historical predictors
  - Mid-range distribution for zero-inflation and gamma distribution of data
- Analysis:
  - Logistic regression alone ($p < 0.05$)
  - GLM gamma + logistic regression ($p < 0.025$ for each test)

MODEL 1:

SAMPLE SIZE MODEL

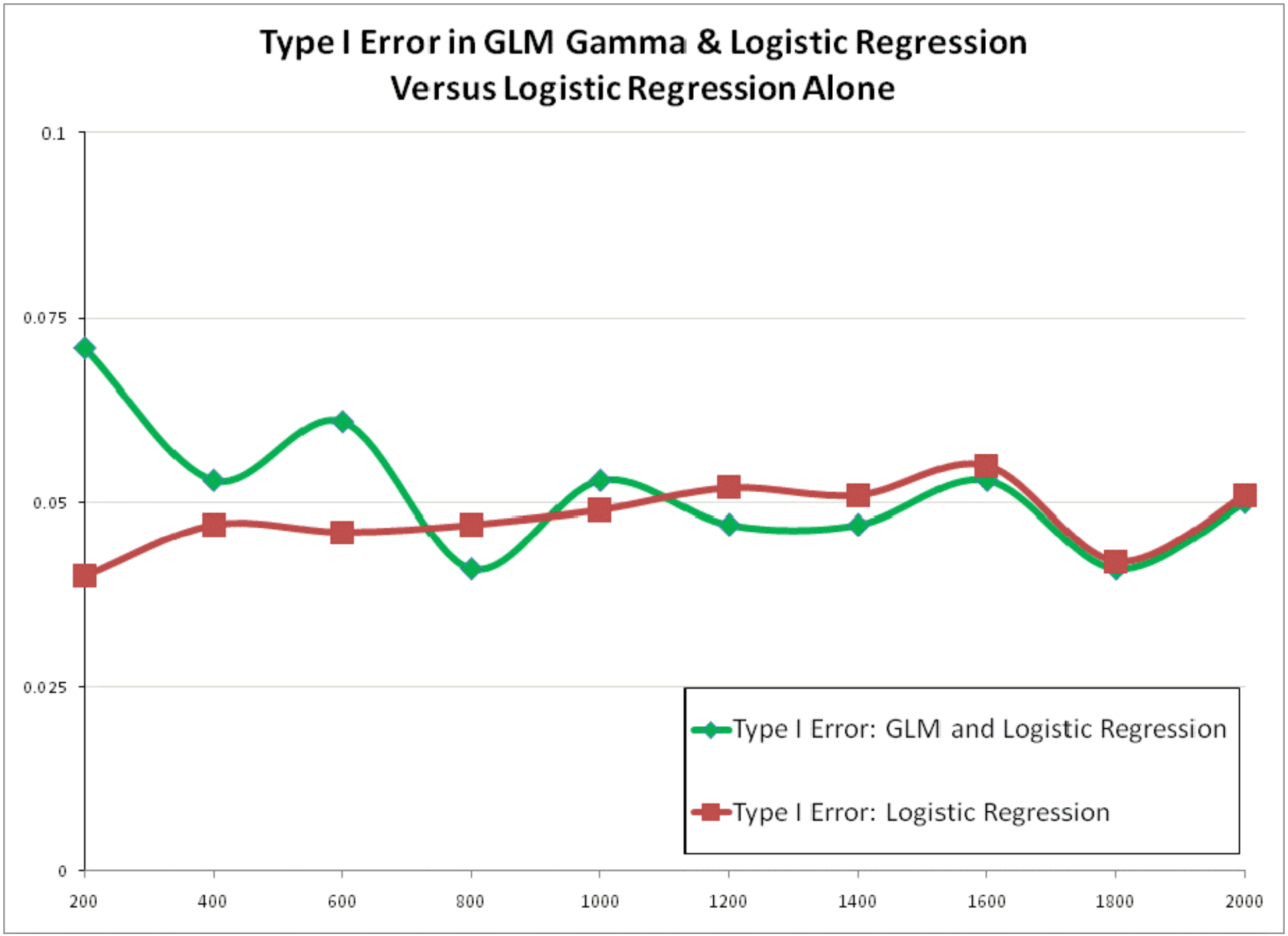
Simulated Model 1

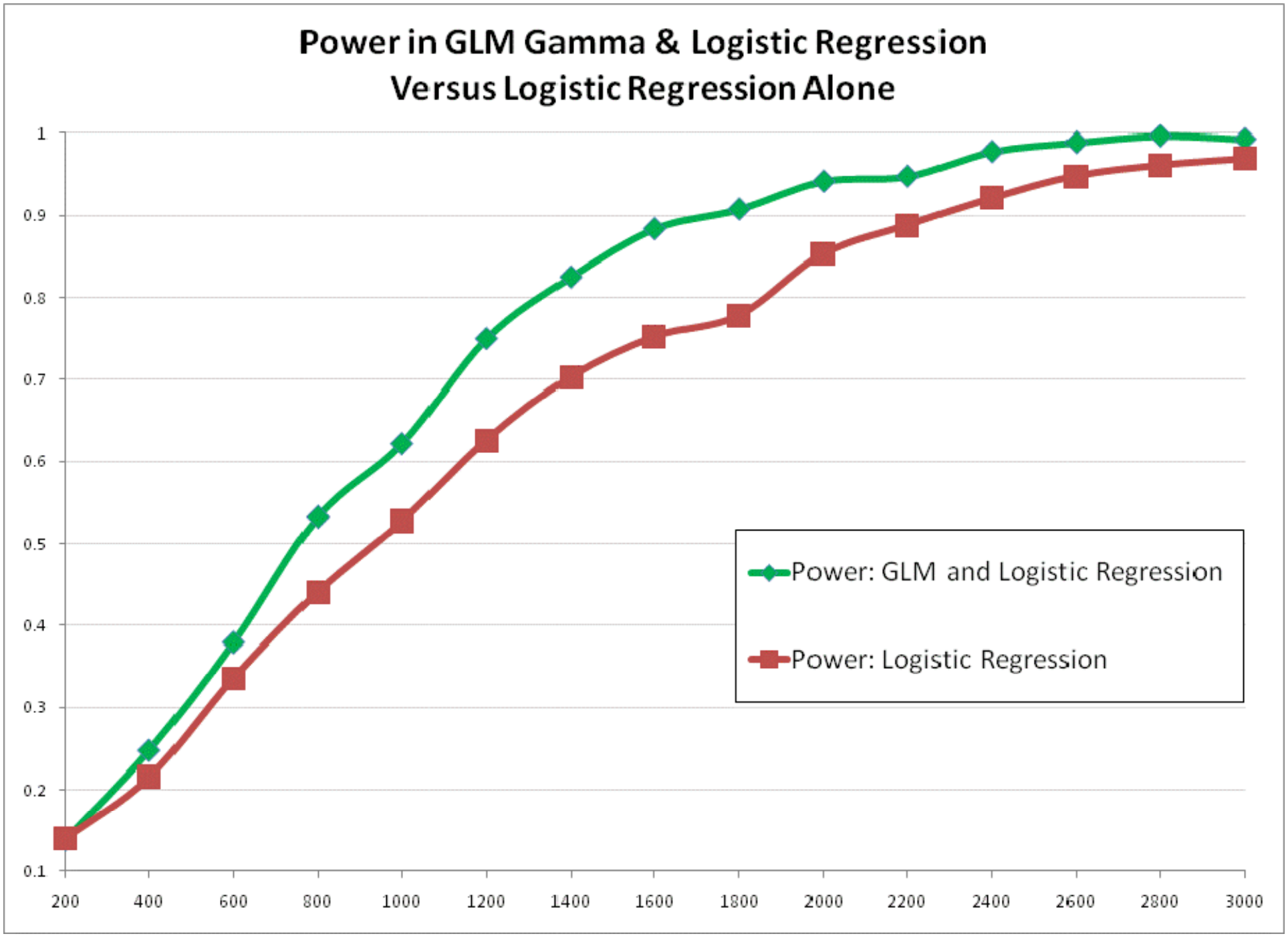
Logistic Regression (n=2850)

Predictor	Frequency (%)	Percent with 0% Non-adherence	Odds Ratio (95% CI)	p-value
0	912 (32)	67.3	Ref	-
1	1311 (46)	65.8	1.07 (0.90, 1.28)	0.44
2	427 (15)	60.2	1.36 (1.07, 1.73)	0.01
3	200 (7)	58.5	1.46 (1.07, 2.00)	0.02

Generalized Linear Model (n=1000)

Predictor	Frequency (%)	Mean % Non-adherence	Coefficient (95% CI)	p-value
0	298 (29.8)	25.7	Ref	-
1	449 (44.9)	27.3	1.06 (0.93, 1.21)	0.37
2	170 (17.0)	31.1	1.21 (1.02, 1.43)	0.03
3	83 (8.3)	32.5	1.26 (1.05, 1.53)	0.02


Type I Error in GLM Gamma & Logistic Regression
Versus Logistic Regression Alone
0.1
0.075
0.05
0.025
0
200
400
600
800
1000
1200
1400
1600
1800
2000
Type I Error: GLM and Logistic Regression
Type I Error: Logistic Regression


Power in GLM Gamma & Logistic Regression
Versus Logistic Regression Alone
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
2600
2800
3000
Power: GLM and Logistic Regression
Power: Logistic Regression

MODEL 2:

GAMMA DISTRIBUTION MODEL

Logistic Regression

Logistic Regression (n=2850)				
Predictor	**Frequency (%)**	**Percent with 0% Non-adherence**	**Odds Ratio (95% CI)**	**p-value**
0	912 (32.0)	67.3	Ref	-
1	1311 (46.0)	65.8	1.07 (0.90, 1.28)	0.44
2	427 (15.0)	60.2	1.36 (1.07, 1.73)	0.01
3	200 (7.0)	58.5	1.46 (1.07, 2.00)	0.02

Close Together

Generalized Linear Model (n=1000)

Predictor	Frequency (%)	Mean % Non-adherence	Coefficient (95% CI)	p-value
0	298 (29.8)	25.7	Ref	-
1	449 (44.9)	24.0	0.93 (0.82, 1.07)	0.32
2	170 (17)	24.1	0.94 (0.78, 1.12)	0.48
3	83 (8.3)	26.5	1.03 (0.83, 1.28)	0.79

Mid-Range

Generalized Linear Model (n=1000)

Predictor	Frequency (%)	Mean % Non-adherence	Coefficient (95% CI)	p-value
0	298 (29.8)	25.7	Ref	-
1	449 (44.9)	27.3	1.06 (0.93, 1.21)	0.37
2	170 (17.0)	31.1	1.21 (1.02, 1.43)	0.03
3	83 (8.3)	32.5	1.26 (1.05, 1.53)	0.02

Spread Apart

Generalized Linear Model (n=1000)

Predictor	Frequency (%)	Mean % Non-adherence	Coefficient (95% CI)	p-value
0	298 (29.8)	25.7	Ref	-
1	449 (44.9)	29.7	1.15 (1.01, 1.31)	0.03
2	170 (17.0)	31.3	1.22 (1.03, 1.44)	0.02
3	83 (8.3)	34.2	1.33 (1.10, 1.61)	0.003

MODEL 3:

ZERO INFLATION MODEL

GLM: Gamma Distribution

Generalized Linear Model (n=1000)				
Predictor	**Frequency (%)**	**Mean % Non-adherence**	**Coefficient (95% CI)**	**p-value**
0	298 (29.8)	25.7	Ref	-
1	449 (44.9)	27.3	1.06 (0.93, 1.21)	0.37
2	170 (17.0)	31.1	1.21 (1.02, 1.43)	0.03
3	83 (8.3)	32.5	1.26 (1.05, 1.53)	0.02

Close Together

Logistic Regression (n=2850)

Predictor	Frequency (%)	Percent with 0% Non-adherence	Odds Ratio (95% CI)	p-value
0	912 (32.0)	67.3	Ref	-
1	1311 (46.0)	69.9	0.89 (0.74, 1.07)	0.20
2	427 (15.0)	63.7	1.17 (0.92, 1.49)	0.19
3	200 (7.0)	62.5	1.24 (0.90, 1.70)	0.19

Mid-Range

Logistic Regression (n=2850)

Predictor	Frequency (%)	Percent with 0% Non-adherence	Odds Ratio (95% CI)	p-value
0	912 (32.0)	67.3	Ref	-
1	1311 (46.0)	65.8	1.07 (0.90, 1.28)	0.44
2	427 (15.0)	60.2	1.36 (1.07, 1.73)	0.01
3	200 (7.0)	58.5	1.46 (1.07, 2.00)	0.02

Spread Apart

Logistic Regression

Predictor	Frequency (%)	Percent with 0% Non-adherence	Odds Ratio (95% CI)	p-value
0	912 (32.0)	67.3	Ref	-
1	1311 (46.0)	61.3	1.30 (1.09, 1.55)	0.004
2	427 (15.0)	51.5	1.98 (1.56, 2.50)	<0.001
3	200 (7.0)	50.5	2.02 (1.48, 2.75)	<0.001

Advantages of GLM Gamma + Logistic Regression

- Uses actual values of all data
- Is statistically a powerful tool
- Has acceptable type I error
- GLM can predict mean non-adherence of individuals with > 0% non-adherence
- Shows where variability in data is coming from (i.e. within dichotomized outcome, within the degree of non-adherence, or both).

Disadvantages of GLM Gamma + Logistic Regression

- Initially may be more complex and technical
  - Assistance of biostatistician is key (which is really an advantage)
- It is a 2-step process
  - No one answer to summarize all data
  - More lengthy interpretation
- Little gain with its use in scenarios with large zero inflation and little spread in gamma distribution

Conclusion

- As it is critical to assess medication adherence using different methods, it is also important to analyze these data by various approaches.
- GLM using a gamma distribution is a powerful tool that can be used in conjunction with logistic regression to get another perspective of the data.

Future Directions:

- Use of this 2-step analysis approach to analyze ARV adherence data assessed using various methods (e.g. pharmacy refill records, pill count, MEMs caps, etc)
- Use of other analysis methods: zinb, zip

Special Thanks To …

- Eric Vittinghoff, PhD
- Torsten Neilands, PhD
- Charles McCulloch, PhD
- Estie Hudes, PhD
- Mallory Johnson, PhD
- Samantha Dilworth, MS

The project described was supported by Award F32MH086323 and U10MH057616 from the **National Institute of Mental Health**.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Institutes of Health.