A Longitudinal Model and Graphic for Benefit-Risk Analysis, With Case Study

Jonathan Norton, Ph.D.
Center for Drug Evaluation and Research
U.S. Food and Drug Administration

Disclaimer

This presentation reflects the views of the author and should not be construed to represent FDA's views or policies.

Benefit-Risk Assessment is Hard

The following comments are from a 2011 Advisory Committee meeting on a first-in-class drug. The committee voted 9-6 against approval:

- *I voted no. I actually like this drug.*
- *I voted no, but it was the closest of calls. I changed my mind about four times in the last 10 seconds.*
- *I voted no. And I agree with all of my colleagues that voted both yes and no… I went back and forth and back and forth.*
- *I voted yes. I have a hard time disagreeing with almost everything that's been articulated here today.*

Why is it hard?

A non-exhaustive list:

- Benefits and harms (risks) are measured on different scales
- Many possible harms
- Missing data
- Benefit-risk tradeoff can change over time
- Benefits and harms may not be independent, e.g., active treatment responders may be at higher risk for adverse event (AE)

A Restricted Case

- Focus on treatments that provide mainly/entirely symptomatic relief, e.g., analgesics
- This simplifies handling of dropouts*. If they felt better on study treatment, would have probably stayed on it. (Non-medical reasons possible, but these are equally likely in all arms.)

*The same logic applies to patients who discontinue study treatment but stay in study, if this option is available.

Original Chuang-Stein model

- Introduced in Chuang-Stein, et al. (1991)*
- Each patient's outcome is one of following:
  1. Benefit, w/out adverse event (AE)**
  2. Benefit + AE
  3. Neither benefit nor AE
  4. AE only
  5. Early withdrawal due to AE
- Advantages: Deals with association between benefit and harm; addresses early discontinuation problem.

* Revised in Chuang-Stein (1994), but focus here on original model
** "Serious side effects"

Revised model

- Count all withdrawals equally (symptomatic relief)
- Consider benefit-risk as dynamic process, allowing patients to change state over course of trial

Application: Chronic Pain

- Mild-to-moderate severity AE's common
- Benefit is subjective
- High discontinuation rates. Traditional missing data methods arguably inapplicable, because nothing is really "missing".
- Might expect benefit-risk profile to change over time as patients develop tolerance for drug (opiate), adjust to side effects, etc.
- Perception that current trial designs are inadequate to show true efficacy (ACTTION initiative)

Case Study #1 Hydromorphone

- Extended-release hydromorphone (HM) tablet approved March 2010. I was the primary stat reviewer.
- Advisory committee considered risks of abuse and diversion. Did ***not*** attempt to model these risks.

Study Design

- 12-week DB trial. Pain recorded daily on 11-point rating scale ("No pain" → "Worst Possible pain").
- Randomized withdrawal design – patients titrate to effective dose, then randomized to active drug or matching placebo
- 58% of subjects were titrated to effective dose and randomized
- Limited amount of rescue medication

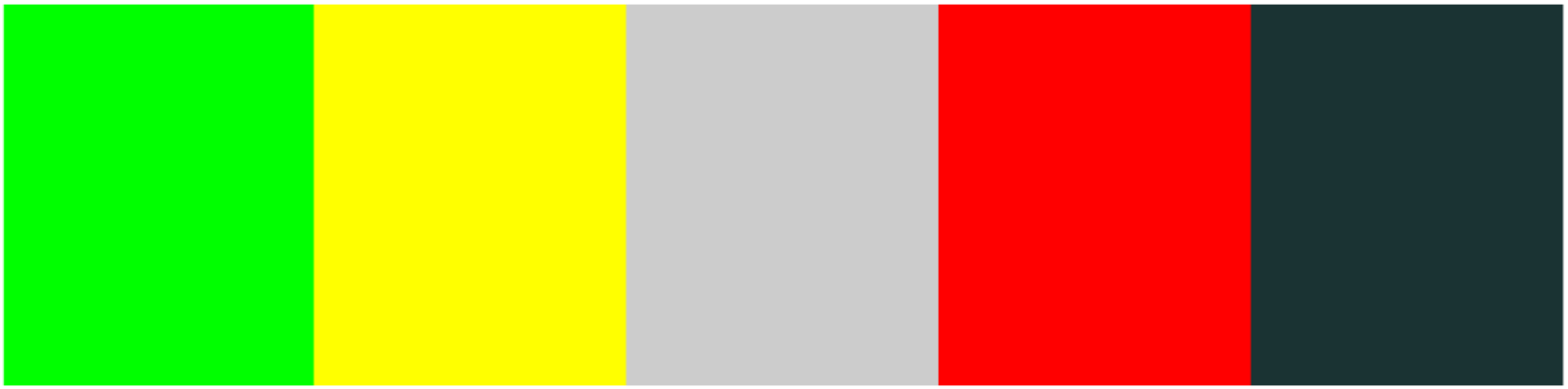
Chuang-Stein model for HM Study

- Benefit = 30% reduction in pain from screening baseline (Farrar et al, 2001)
- AE = moderate-to-severe adverse event in DB phase of study

Individual Response Profile

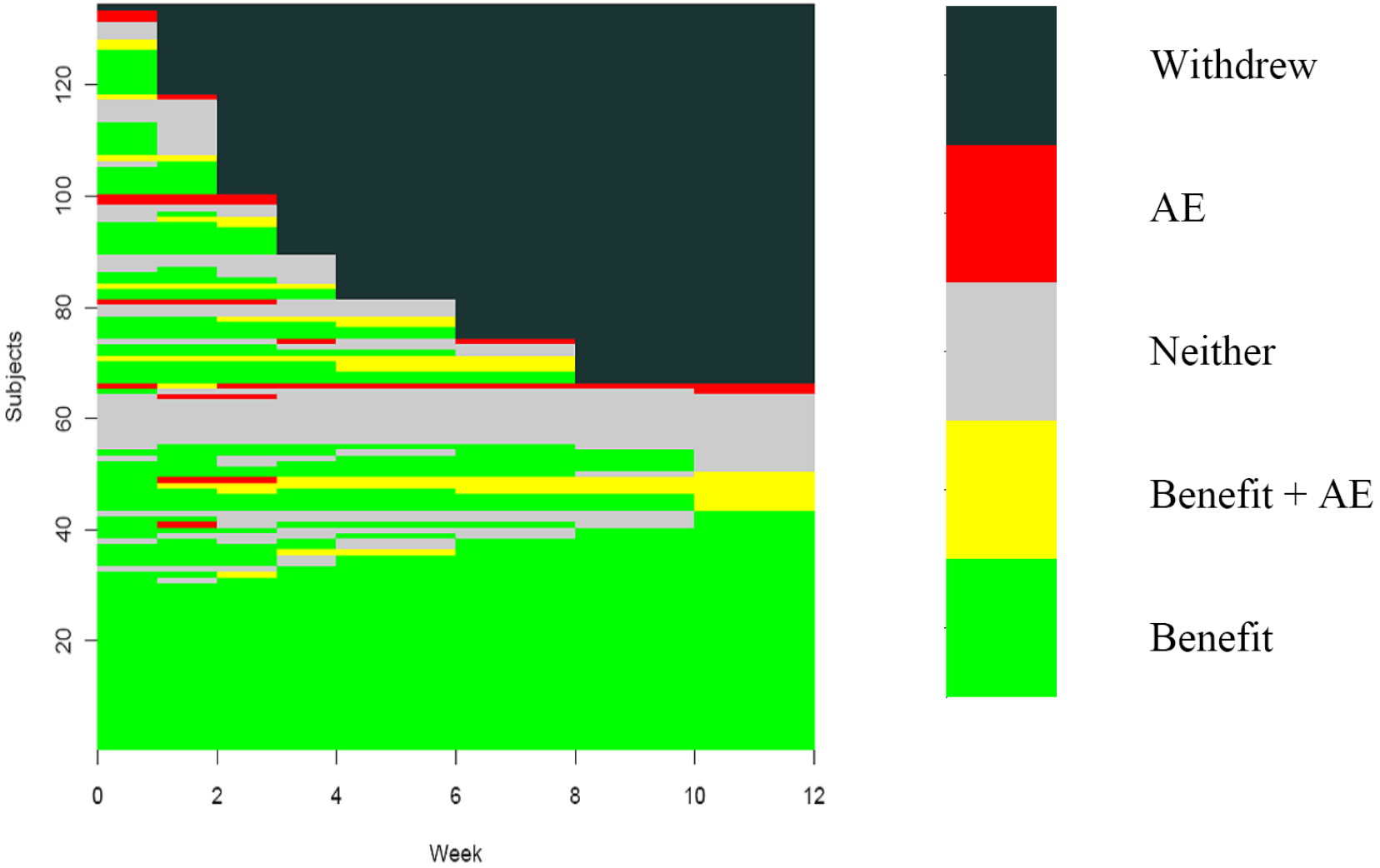
- IRP graphic has **one row for each subject**
- X-axis is time
- Colors used to distinguish individual states
- Sort rows for clarity. Preferred method uses last period as primary sort key, then second-to-last period, etc.
- Could be used for any longitudinal, categorical study outcome, not just Chuang-Stein model
- Similar graphics: event history (Dubin et al, 2001), safety at FDA (Szarfman et al, 1997), lasagna plots (Swihart et al, 2010)

Legend

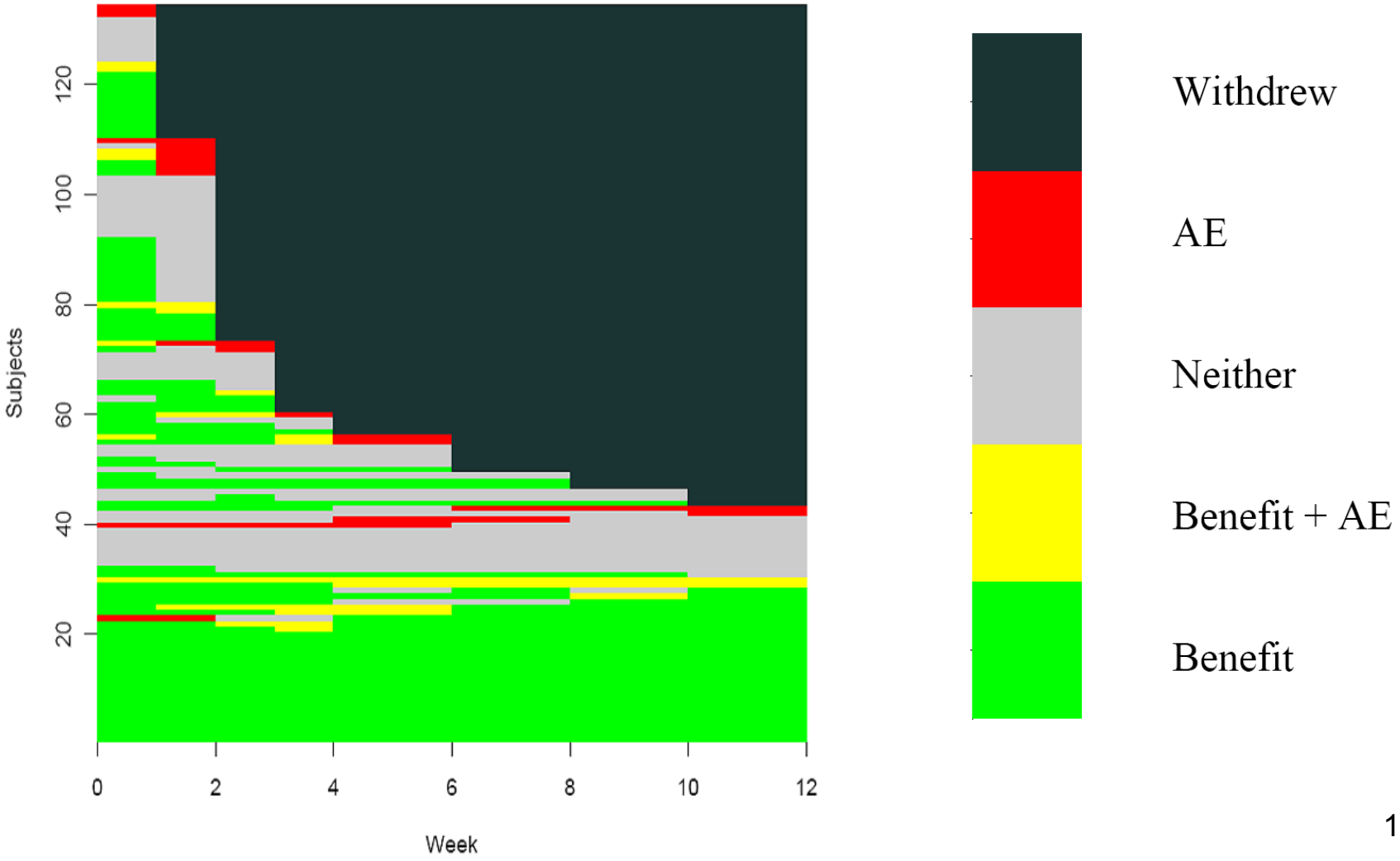


Note: Color-blind accessible figures start on slide 38.

HM Results

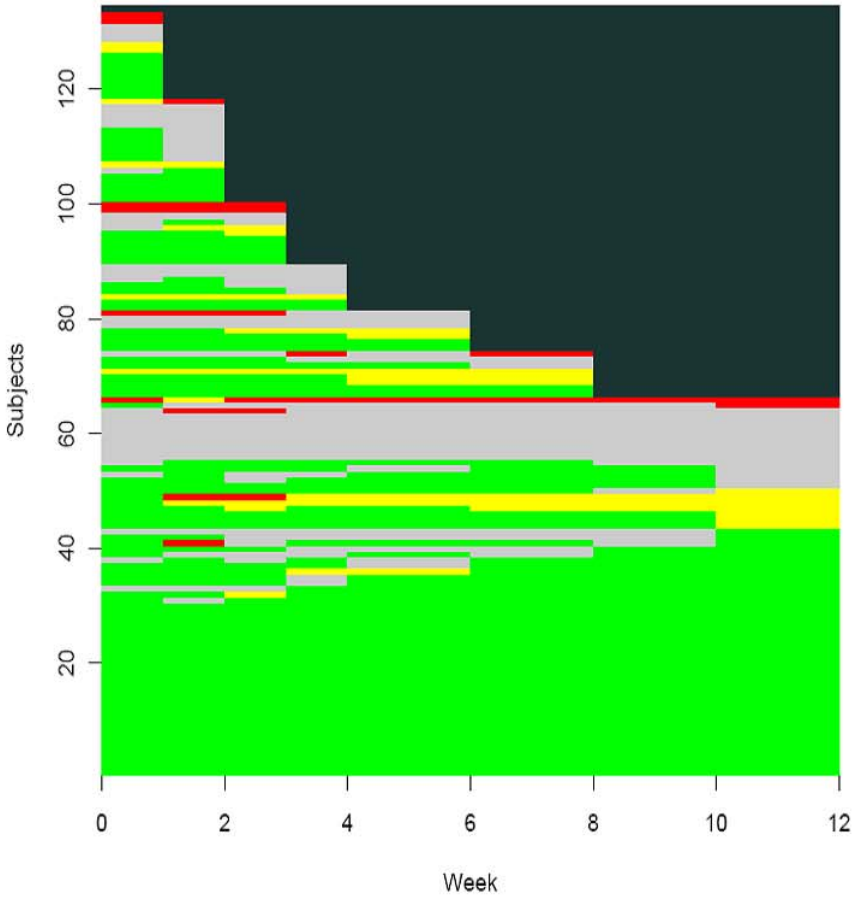

Subjects
20
40
60
80
100
120
0
2
4
6
8
10
12
Week
Withdrew
AE
Neither
Benefit + AE
Benefit

Placebo Results

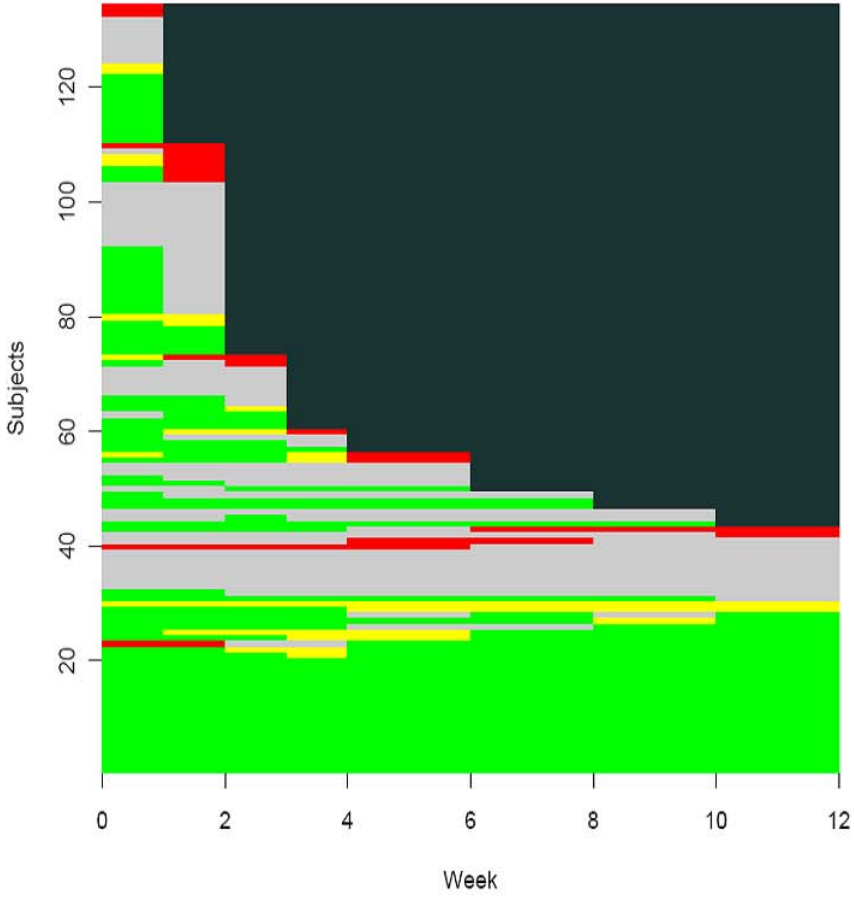


Comparison of IRPs

HM



Placebo



Additional Case Studies

Joint with Sanatan Saraf, FDA Intern

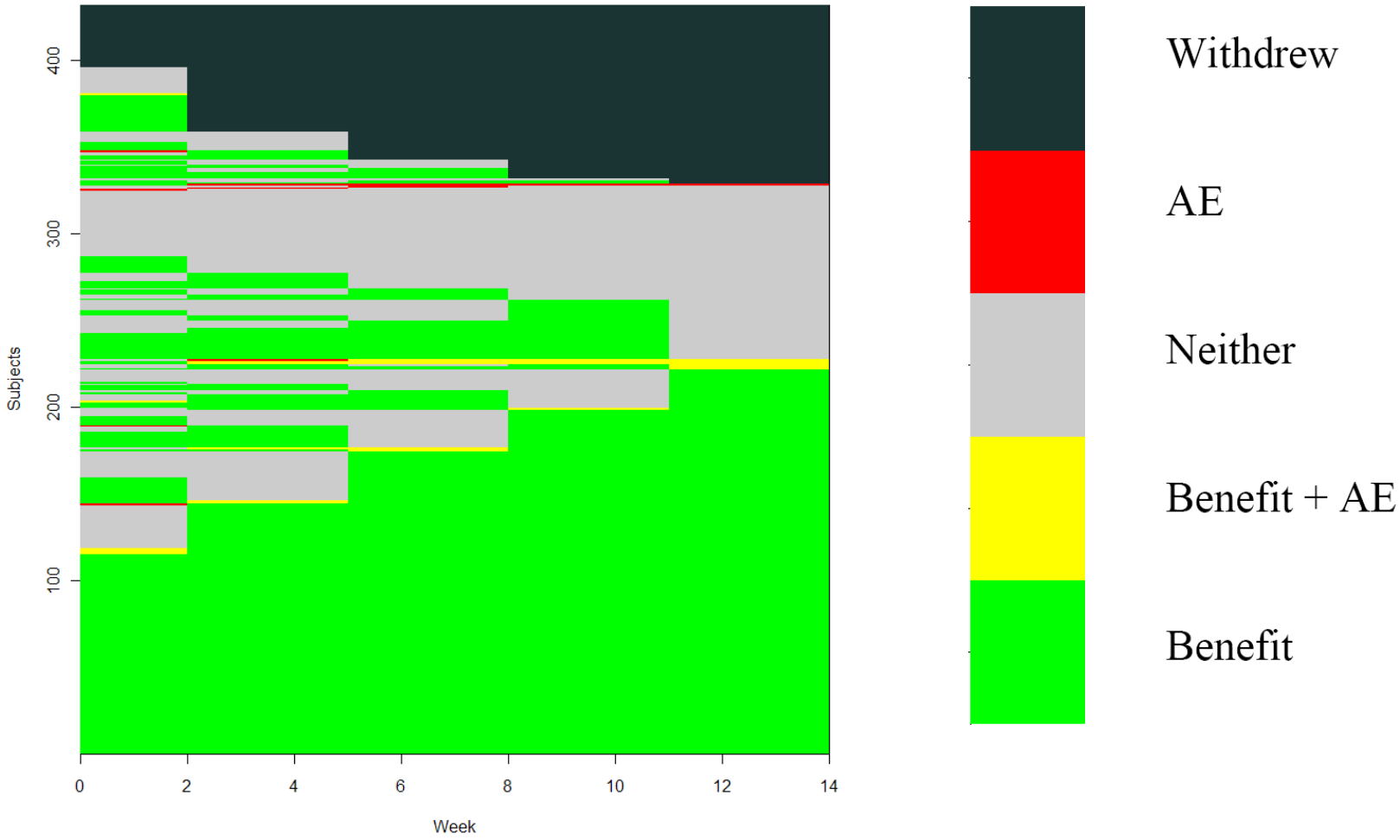
Case Study #2: Tramadol

- Extended-release tramadol approved in 2008
- Weak opiate with other modes of action. Genetic polymorphism (CYP2D6) affects metabolism.
- Design – also randomized withdrawal:
  - 4 wk OL: 2 wk run-in, 2 wk taper and washout
  - 14 wk DB: 2 wk titration, then 12 wk maintenance
  - 63% of treated patients were randomized

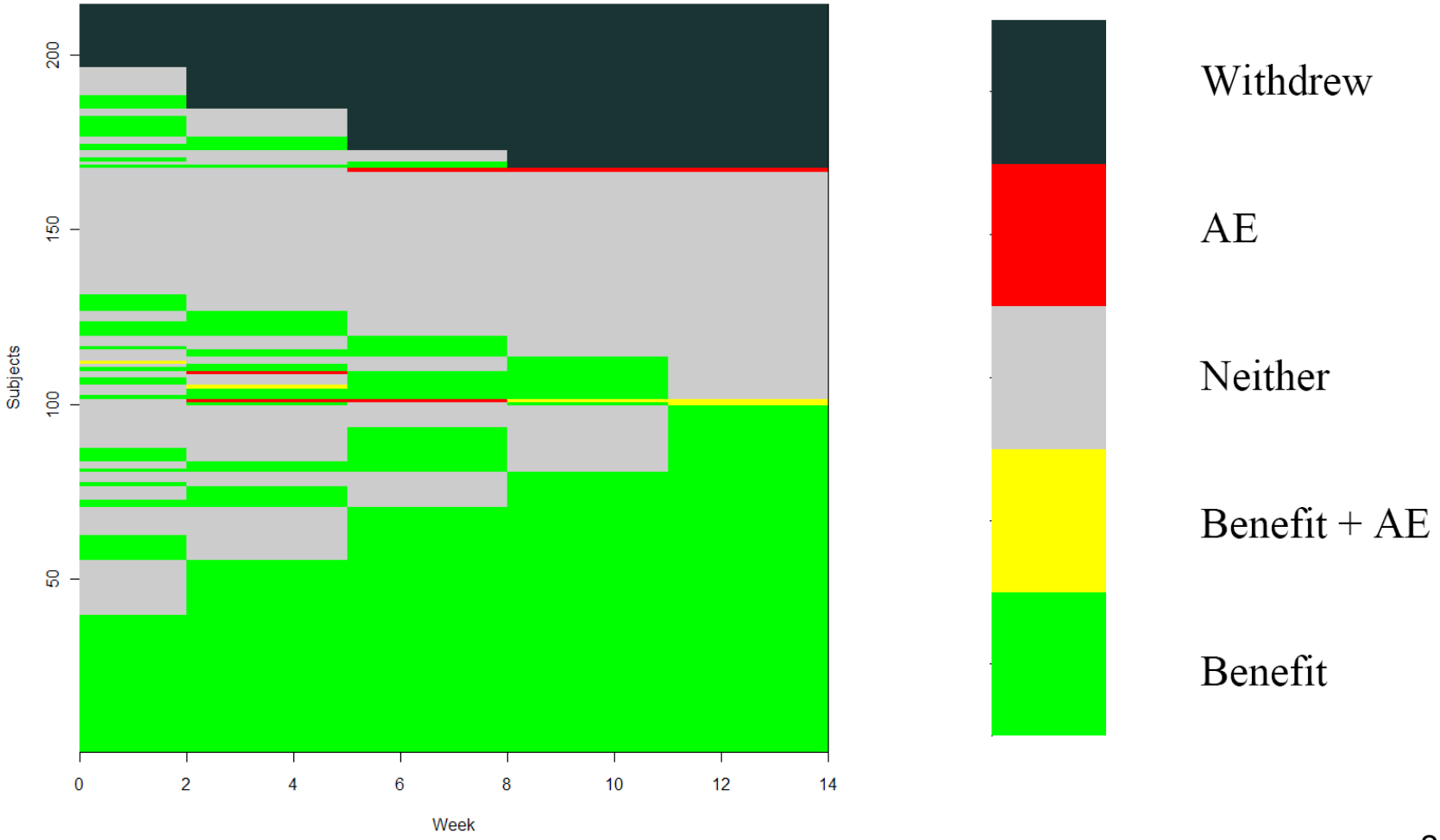
Tramadol (cont.)

- Benefit: 30% reduction in pain from end of washout
- Risk: All severe or serious AEs
  - Onset must be in DB period
  - Broader inclusion criteria for AE used in manuscript in development
- Not “apples to apples” comparison with HM. Drugs would essentially have to be on same protocol for fair comparison.

Tramadol

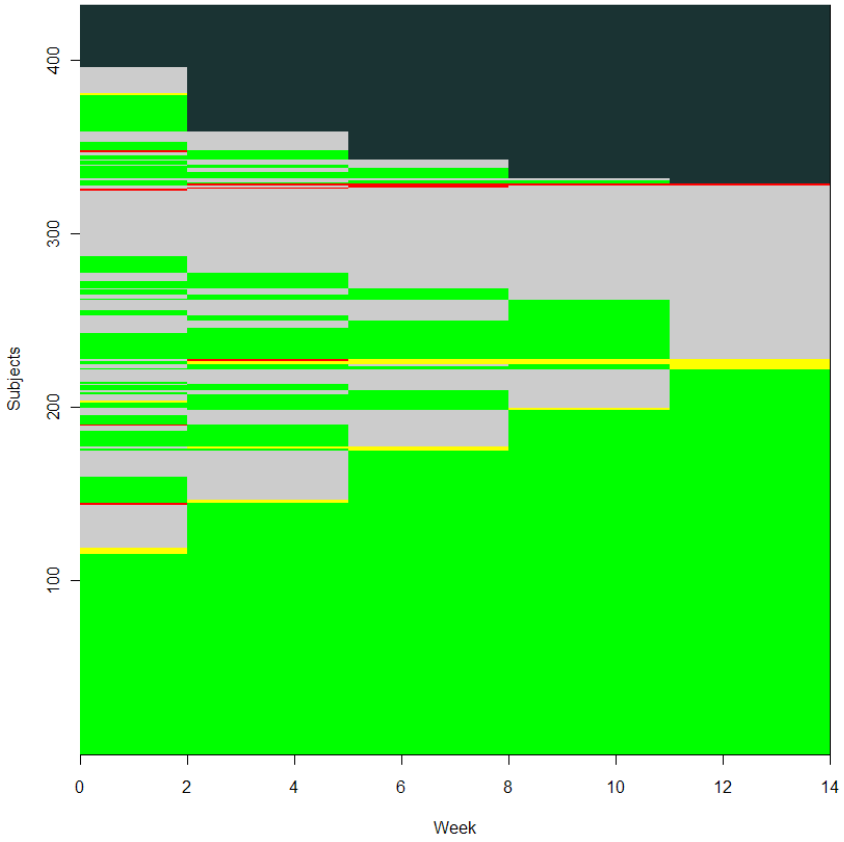

Subjects
100
200
300
400
0
2
4
6
8
10
12
14
Week
Withdrew
AE
Neither
Benefit + AE
Benefit

Placebo

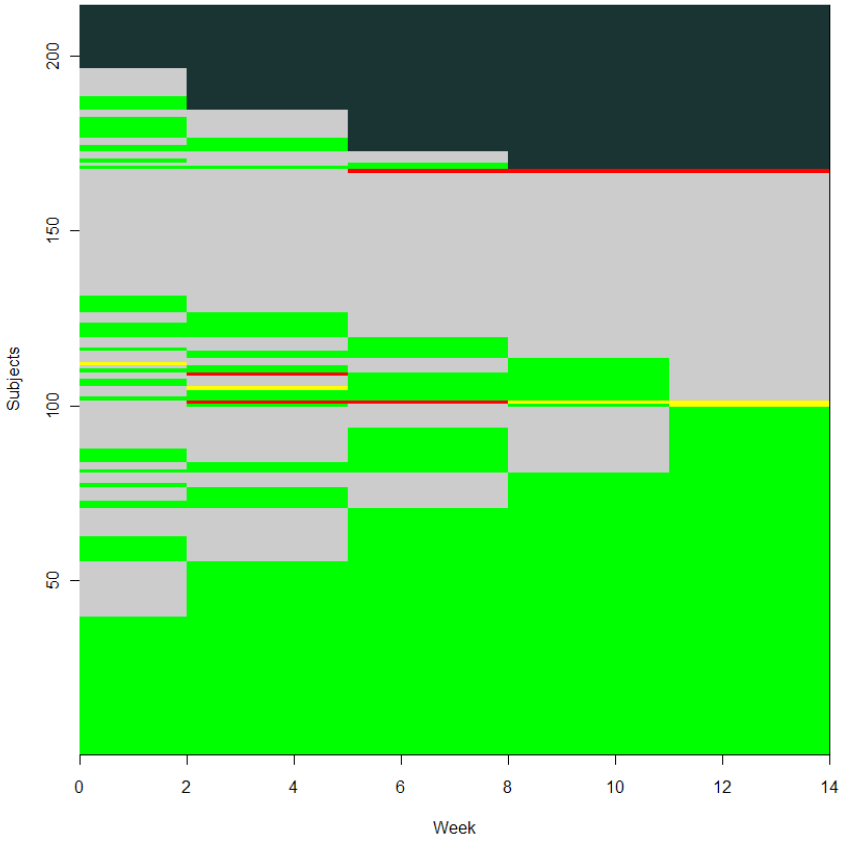

Subjects
200
150
100
50
0
2
4
6
8
10
12
14
Week
Withdrew
AE
Neither
Benefit + AE
Benefit

Comparison

Tramadol



Placebo



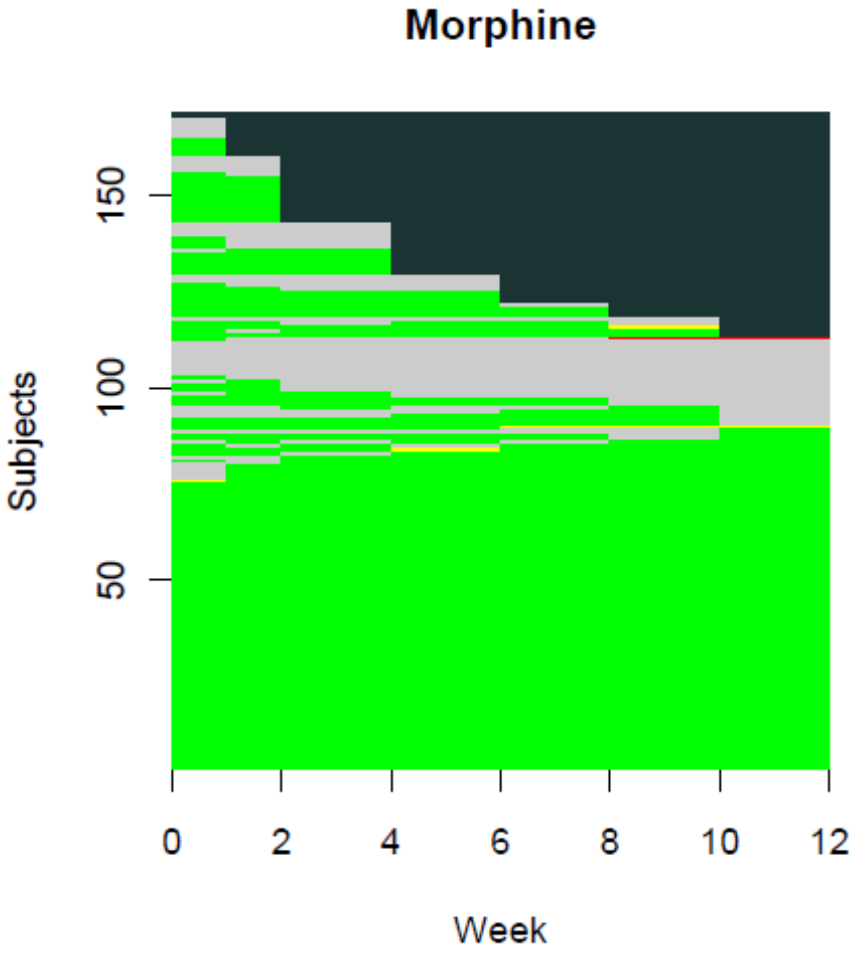
Case Study #3: Morphine

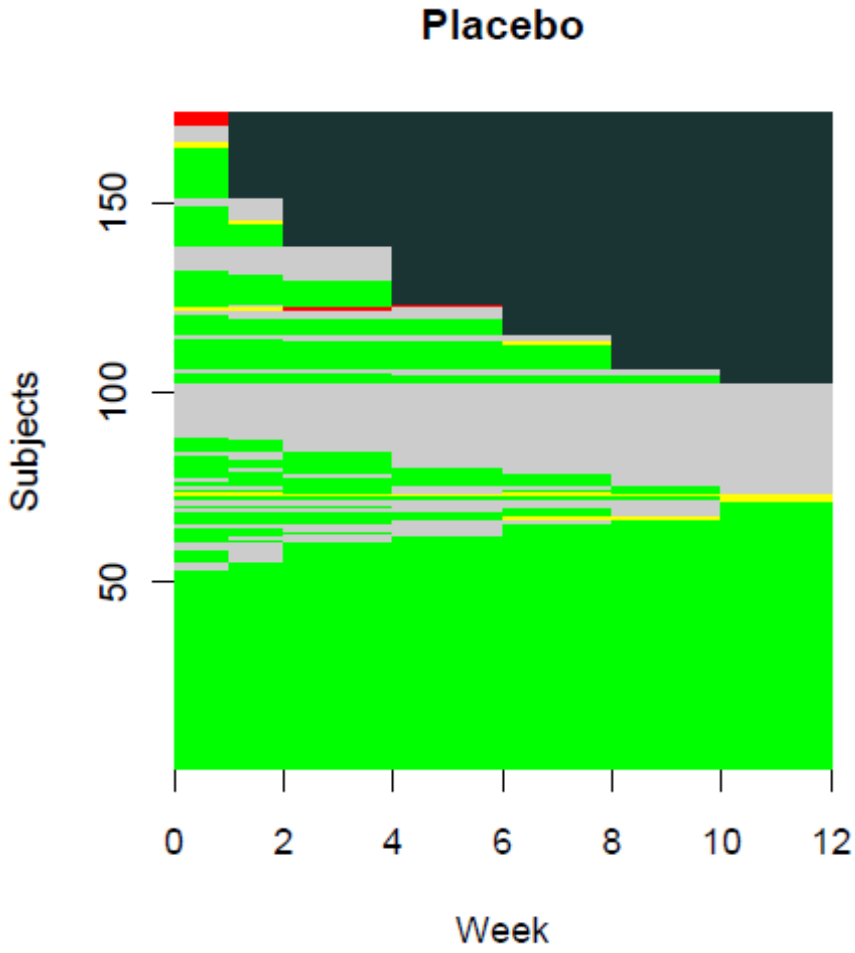
- Extended-release morphine approved in 2009
- Design – also randomized withdrawal:
  - 9 wk OL: 2 wk screening, 1 wk washout and 6 wk Titration period
  - 12 wk DB: 12 wk maintenance
  - 62.9% (344/547) of patients who entered titration period were randomized

Morphine (cont.)

- Same definitions of Benefit and Risk as tramadol
- Again, not “apples to apples” comparison with other cases

Comparison


Morphine
Subjects
150
100
50
0
2
4
6
8
10
12
Week


Placebo
Subjects
150
100
50
0
2
4
6
8
10
12
Week

Testing

- Temporal correlations
- Hard to avoid assigning score to each category and weight to each time period
- One approach: Weighted sum of *signed* Pearson correlation coefficients, with *r* the correlation between the treatment indicator and the outcome for a period. Use permutation test.

Testing (cont.)

$$S = \sum_{t=1}^{T} w_t Y_t$$

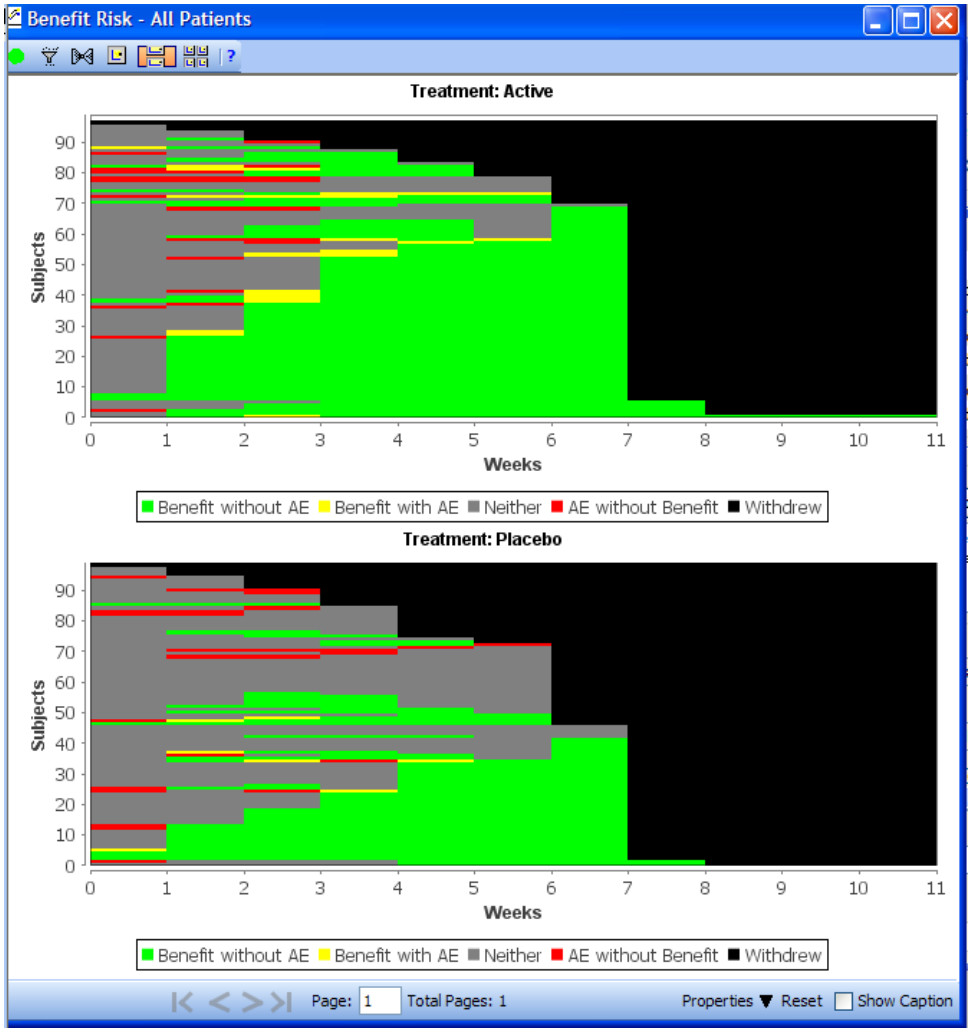
where *S* is test statistic, *w* weight on period t, *Y* is score for outcome on period *t*

- Ad-hoc approach:
  - o Equal spacing for categories, e.g., $Y \in \{1,\ldots,5\}$
  - o Linearly increase weight over course of trial
- Applying ad-hoc test to HM, for example, superior to placebo with *p* = .0011 (permutation)

Software Implementation

- IRP feature in development for JReview software (Integrated Clinical Systems, Inc.)*

**Figure courtesy of ICS. No financial relationship with speaker.*

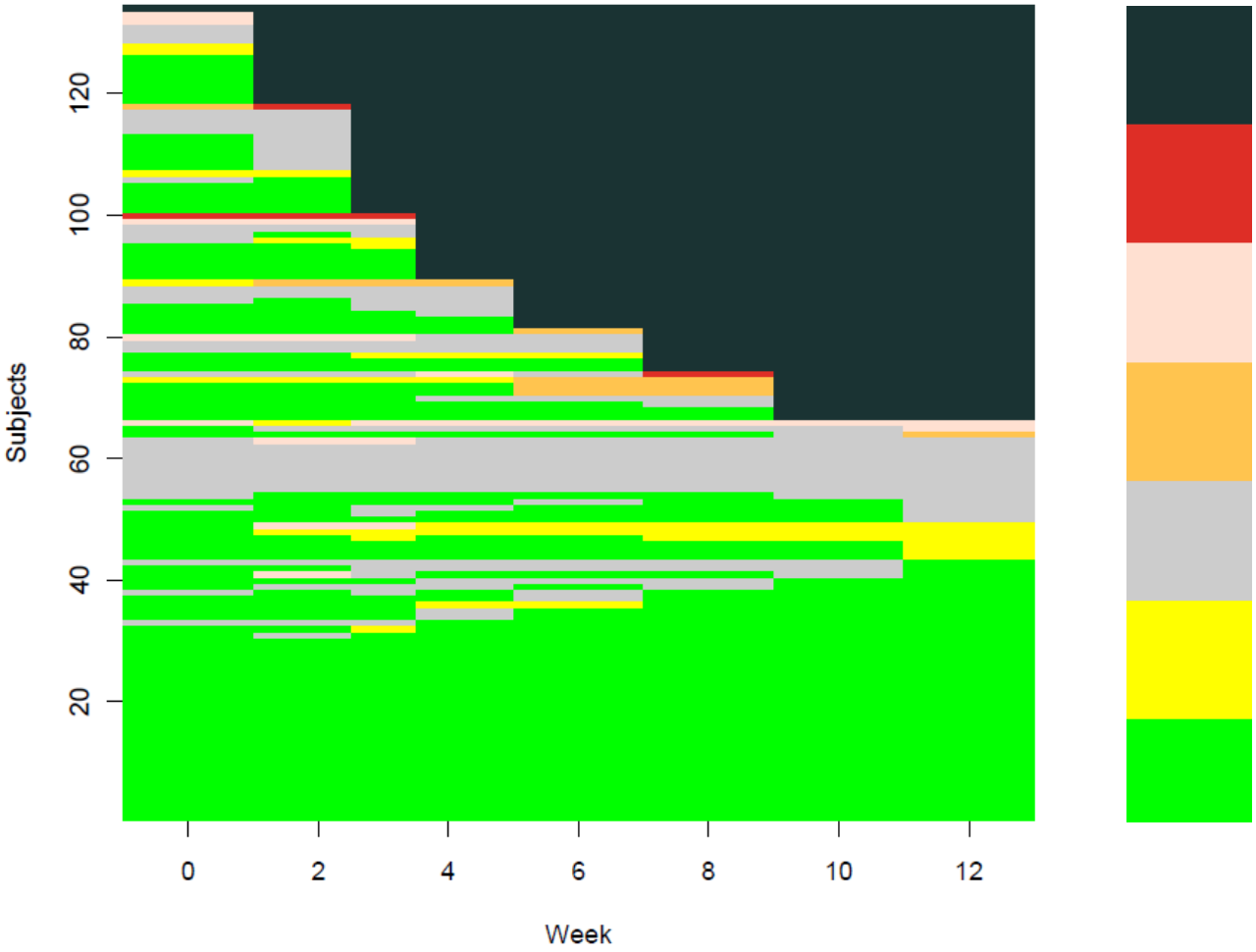


More Colors?

- Frequent suggestion is to use more colors to indicate the severity of an AE
- My preference is for simplicity
- However, here is one possible scheme:

Benefit	Benefit + Mod. AE	Neither	Benefit + Sev. AE	Mod. AE	Sev. AE	Withdrew

HM – With severity



Concluding Opinions

- The human mind can only handle a certain amount of information. Tools like the IRP can help to reduce a complex problem to manageable chunks.
- Statisticians, clinicians, psychologists, and other should collaborate on developing these tools. Many dissertations could be written.
- Transparency in decision making is ultimately beneficial to everyone.

Acknowledgments

- Sanatan Saraf (FDA intern, also University of Maryland, Baltimore County)
- Frank Pucino, Pharm.D. (FDA)
- Tom Permutt, Ph.D. (FDA)
- Ana Szarfman, M.D., Ph.D. (FDA)
- FDA for Regulatory Science And Review Enhancement grant
- Drug sponsors for providing data

References

- Chuang-Stein, Mohberg, Sinkula (1991). *Stat Med.* 10:1349-1359.
- Chuang-Stein (1994). *Contr Clin Trials*. 15:30-43.
- Dubin, Muller, Wang (2001). *Stat Med.* 20:2951-2964.
- Farrar et al (2001). *Pain*. 94:149-158.
- Swihart et al (2010). *Epidemiology*. 21:621-625.
- Szarfman, Talarico, Levine (1997). *Comprehensive Toxicology.* 4:363-79. Ed: Sipes, McQueen, Gandolfi.

Additional Slides

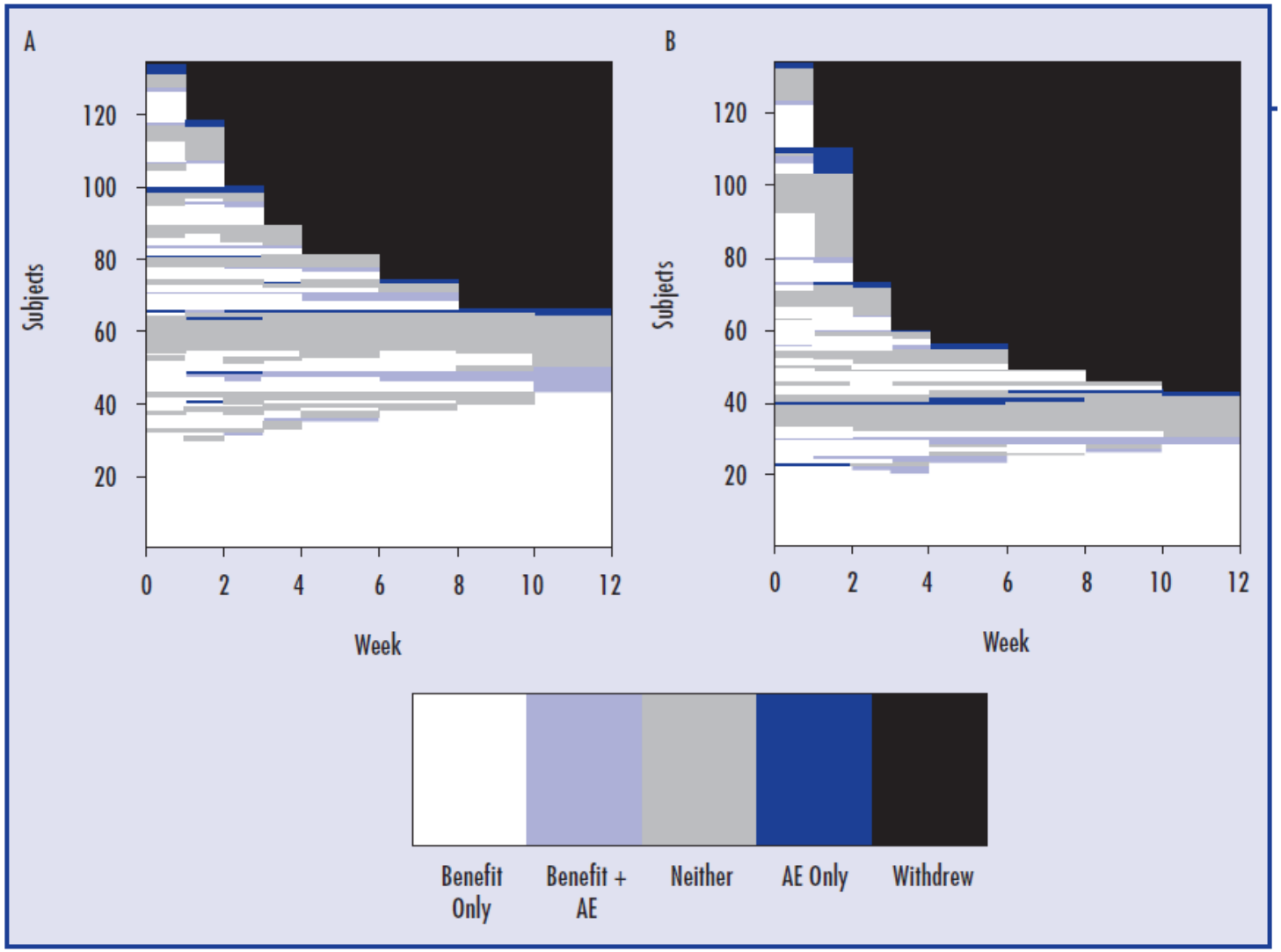
Sample Code

Plotting in R

```
#Notes: Tested in R 2.10.1. Use at your own risk.
library(graphics)
#Toy example with three 4-week time periods and two subjects
Weeks <- c(0,4,8,12)
#Subject 1 was in yellow state, then gray, then red
Subj1 <- c(2,3,4)
#Subject 2 was in green state, then yellow, then withdrew
Subj2 <- c(1,2,5)
Outcomes <- as.matrix(rbind(Subj1, Subj2))
#Sort matrix, using final period as primary key, then using second-to-last period, etc.
SortedOutcomes <- Outcomes[order(Outcomes[,3],Outcomes[,2],Outcomes[,1]),]
nsub <- dim(Outcomes)[1]
stoplightcol <- rgb(c(0,1,.8,1,.1),c(1,1,.8,0,.2),c(0,0,.8,0,.2))
image(Weeks,1:nsub,t(SortedOutcomes),col=stoplightcol)
#accessible purple-gray scheme
accessiblecol <- rgb(c(.75,1,.5,.5,0),c(.75,0,.5,0,0),c(.75,1,.5,.5,0))
image(Weeks,1:nsub,t(SortedOutcomes),col=accessiblecol)
```

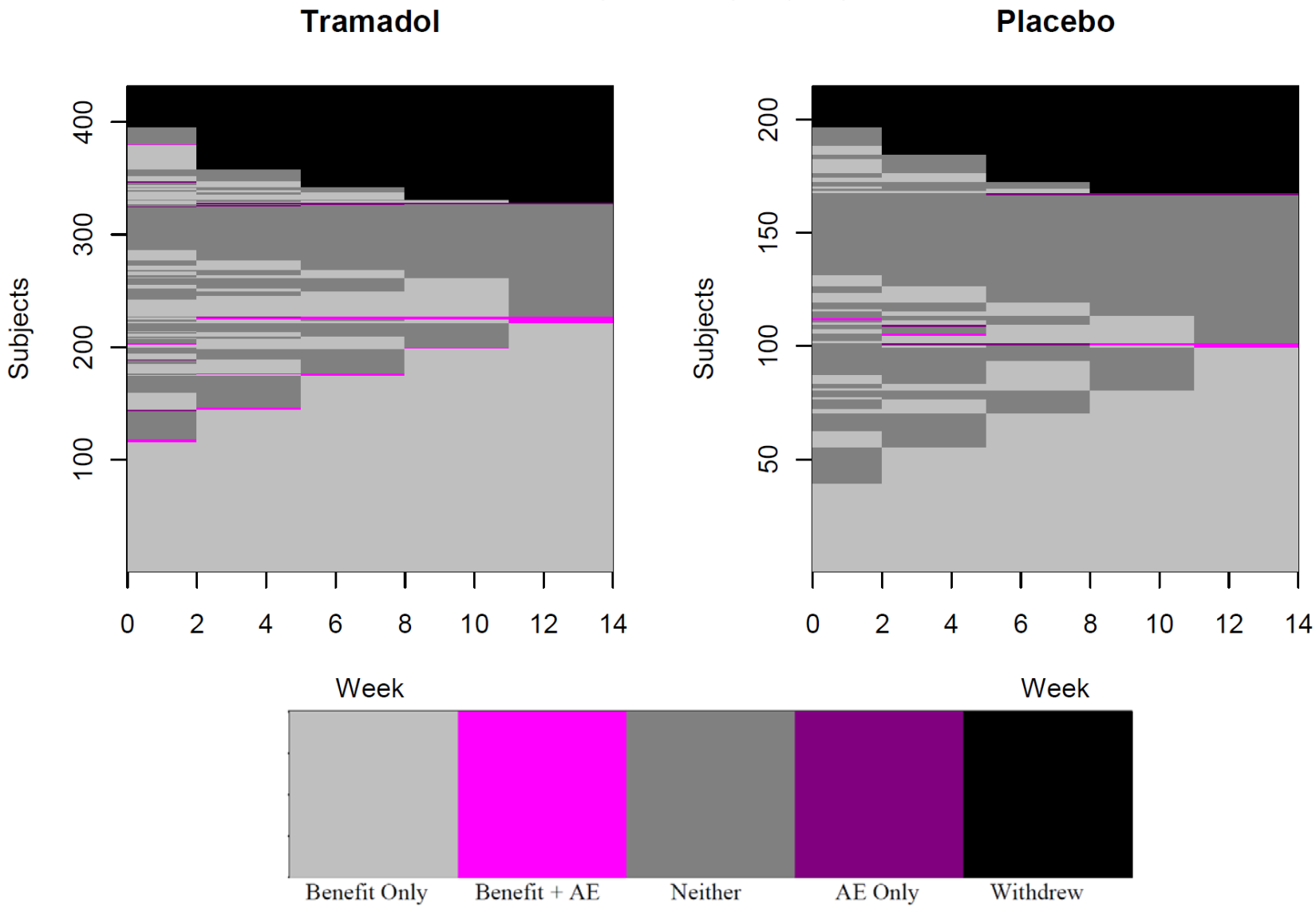
Color-Blind Accessible Figures

Hydromorphone

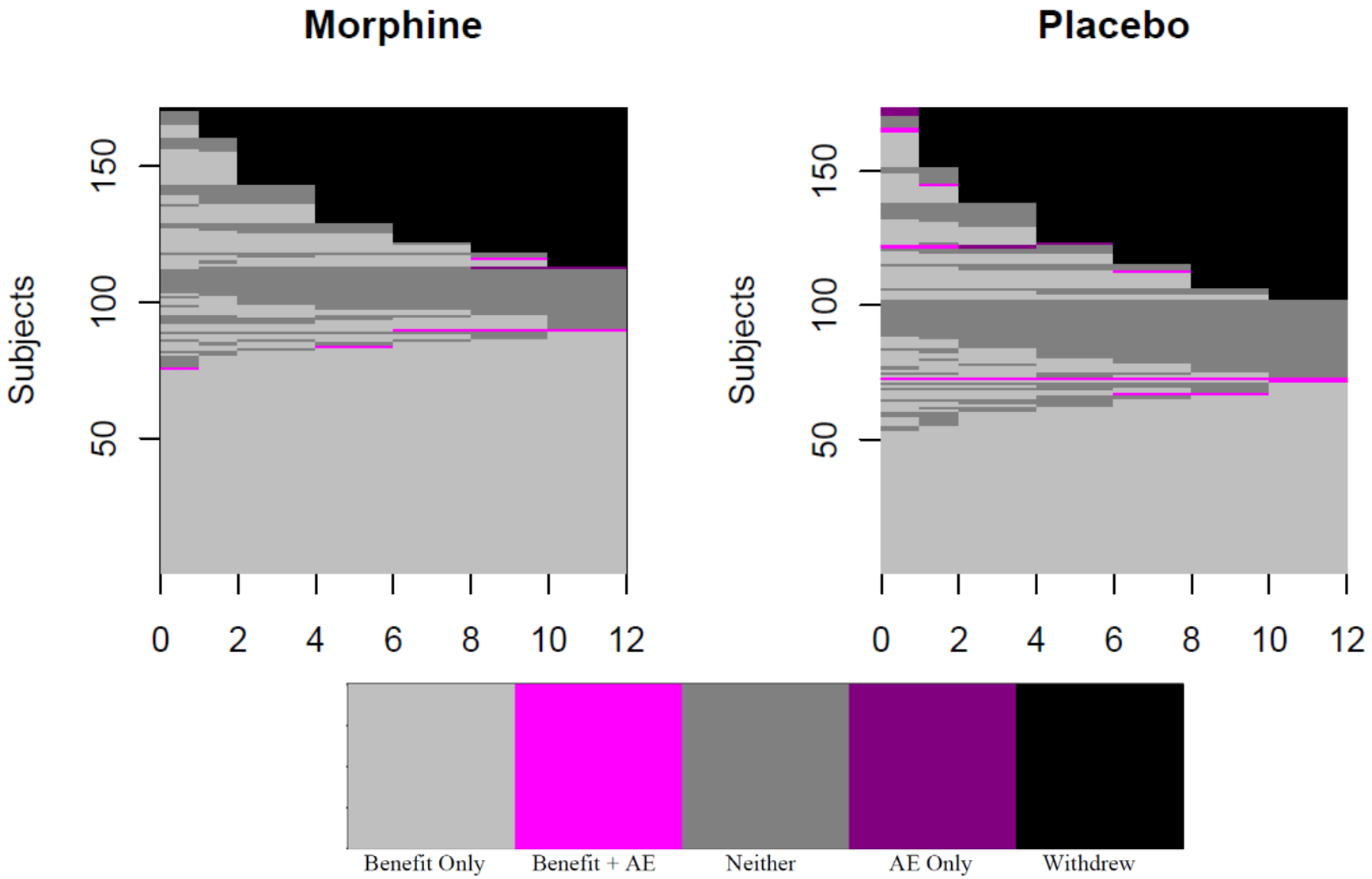


A: HM
B: Placebo

Tramadol


Tramadol
Placebo
Subjects
100
200
300
400
50
150
0
2
4
6
8
10
12
14
Week
Benefit Only
Benefit + AE
Neither
AE Only
Withdrew

Morphine


Morphine
Placebo
Subjects
150
100
50
0
2
4
6
8
10
12
Benefit Only
Benefit + AE
Neither
AE Only
Withdrew