Non-Inferiority and Adaptive Design for HIV Clinical Trials
A Discussion

Greg Soon, Ph.D.
Office of Biostatistics/OTS/CDER/FDA

Emerging Issues in HIV Clinical Trials for new ARVs Roundtable

Washington DC

September 30, 2010

1608 Rhode island ave NW - University of California (across from Double Tree Hotel - far side of circle) (Paid parking in lot next door)

Disclaimer

Views expressed here are of the presenter and not necessarily of the FDA

Non-Inferiority Trial: Where does it Fit?

1. **Superiority Trials** are self-contained: it has all information needed to calculate the new drug's contribution over the control to ensure the new drug is better than placebo
   - Separates the signal from background noise
   - Truly controls Type I error
   - Sloppiness in trial conduct tend to reduce the effect size therefore encourage better trial conduct
   - Independence of evidence from trial to trial
2. **Non-inferiority Trials** has to borrow control vs. placebo info from historical trials
   - "Constant effect assumption," or at least a fraction of it.
   - If control vs. placebo effect is over-estimated by chance, then the inflated margin will lead to wrong decisions for all future trials with the same control
   - If the new trial differs from the historical trials on some factors, factors that affect the response similarly in all arms (main effect) will be cancelled. However interactions will not be cancelled
   - Type I control is always conditional on that we get the control effects right
     - Dependence of trials that share the control either directly or indirectly
   - Sloppiness in trial conduct tend to mask the ineffectiveness of the new drug therefore may encourage poor trial conduct.
   - Comparative efficacy and safety
3. **Epidemiology studies** has to borrow placebo response rate from historical trials
   - "Constant response assumption"
   - Both main effects and interactions matters

How Much Discount in Setting the Margin? Why It Is a Review Issue?

1. When the same trial is repeated, the results will not replicate exactly. The control effect we observed in historical trial could be an over-estimate (50/50). Such an over-estimate can lead to incorrect decisions for the future drug approval, either directly or indirectly. Therefore even if we can replicate the historical trial setting in the new trial there is still a need for discount. The calculation of M1 used the 95% LB as a way of discounting.
2. If we are 100% sure that the new trial is identical to the historical trials then no further discount is needed.
3. In reality this is impossible, especially in HIV treatment where differences in viral mutations, background drug use, patient management criteria, demographical variables, endpoint determination (eg. new assay), missing data and quality of trial, will occur and evolve over time.
4. Margins agreed with FDA during IND are really the best guesses. Often 50% is used as a starting point. Final margin will be a review issue depending on the points discussed above.
5. Clinical margin represents a trade off on the potential loss of efficacy vs. potential gain in safety, tolerability and compliances. Therefore is a review issues as well when the safety, tolerability and compliances are understood.
6. Benefit/Risk assessment

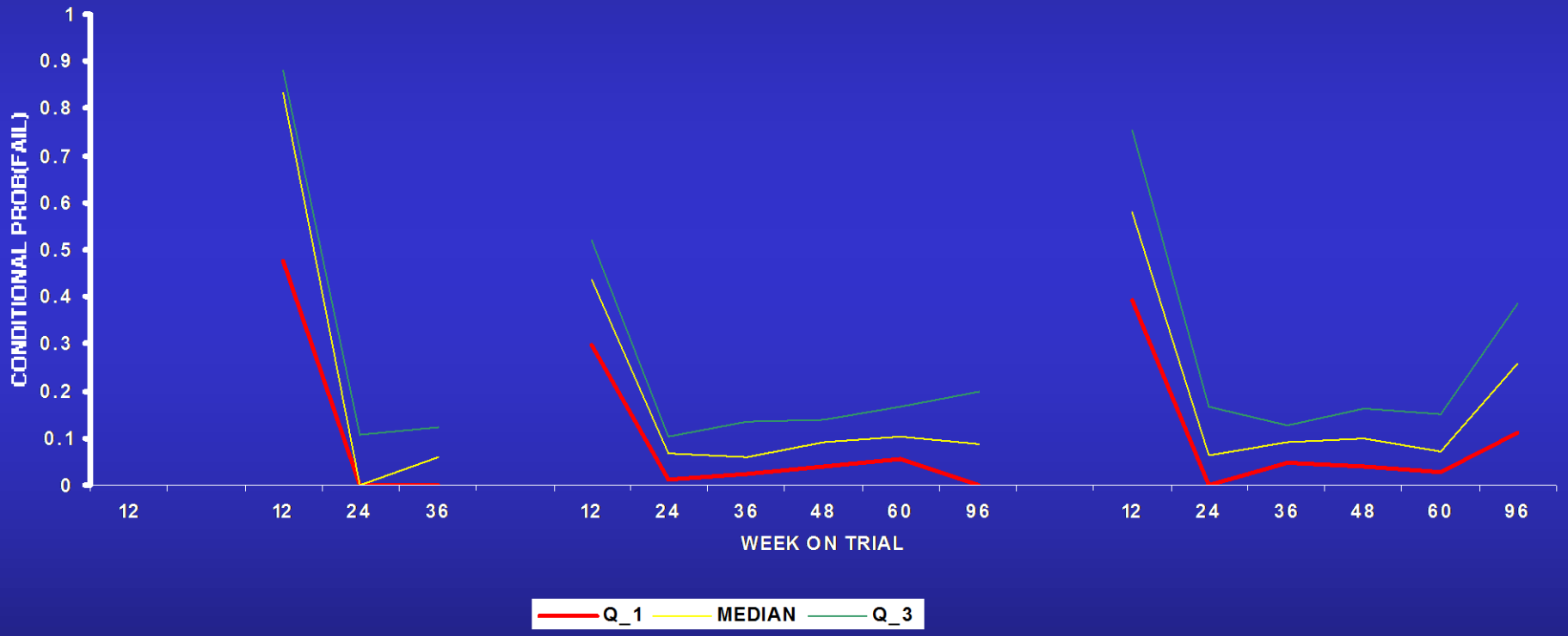
Maintenance Study May Not Be a Good Source for Efficacy Evidence

1. Maintenance studies are usually NI studies
2. Patients are already stable responders and contribution from the control drug is usually not well defined due to lack of placebo controlled trials in this population
   - Less potent regimens could potentially yield good response
   - May not be able to differentiate the drug benefit from placebo had a placebo controlled trial is done
3. Does losing 10% response from a 95% control response rate the same as losing 10% from 70% control response?

Maintenance Study Results May Not Be Generalizable to Non Maintenance Studies

1. The design is advantageous to the test drug for efficacy because control has been used for a while and test drug is still fresh, even if the two drugs have the same potency when started on equal footing
2. The design could be disadvantageous to the test drug for safety because patients intolerant to the control may have dropped out and test drug is still fresh, even if the two drugs have the same magnitude of safety problems

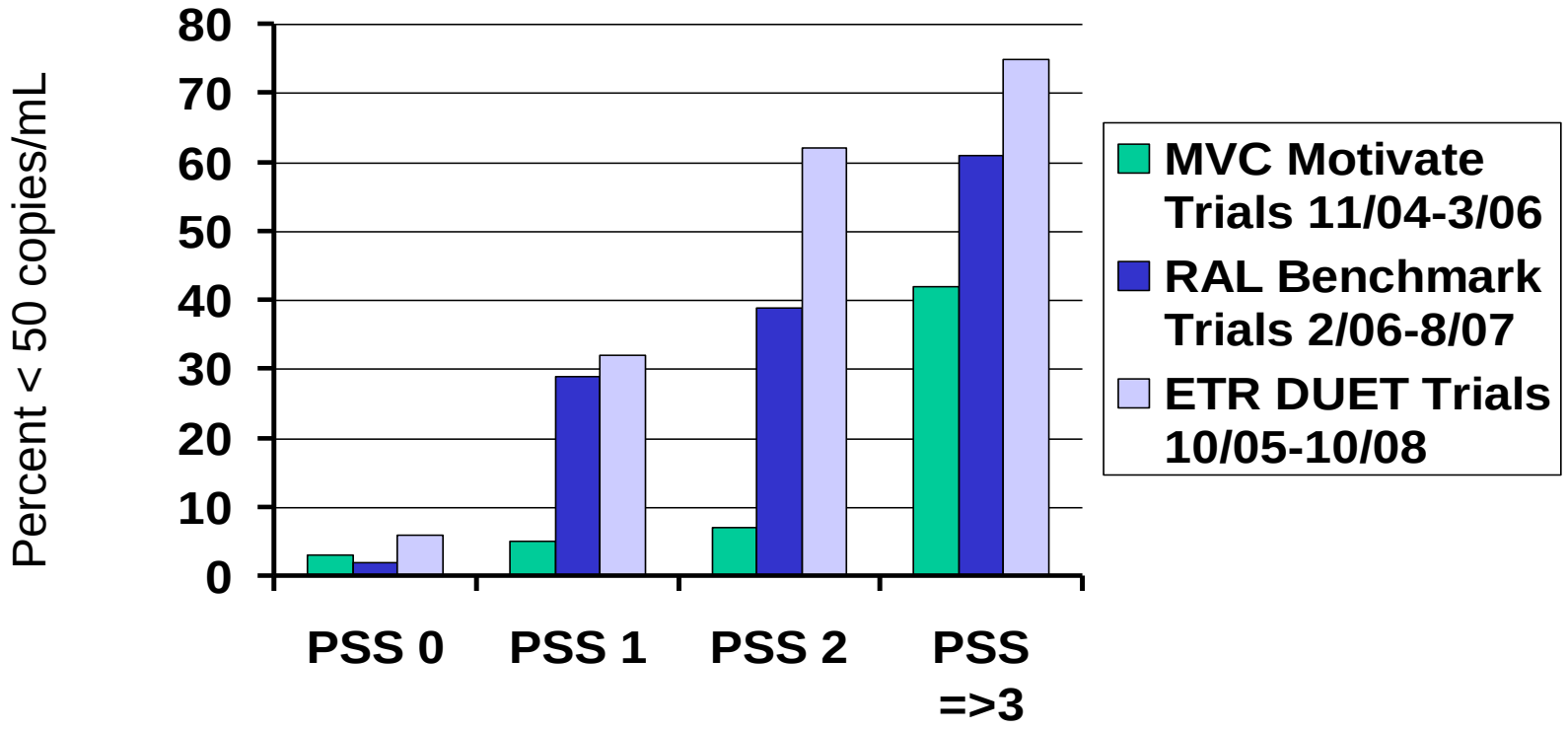
Maintenance Study Results May Not Be Generalizable to Non-Maintenance Studies



Hybrid Design Combining Superiority and NI

1. Fundamental Issue: NI design more difficult with improved OBT response and new drugs introduced in OBT
2. But with good OBT response we could do without the added control
3. A New Proposal - Hybrid Design:
   - Superiority for PSS>=2:
     New drug + OBT vs. placebo + OBT
   - NI for PSS<2:
     New drug + OBT vs. Control + OBT
   - Final analysis will pool evidence from both the superiority and NI parts

Virologic Response (HIV-RNA < 50) for OBT over Trials/Time



Adaptive Design

1. Adaptive designs in anti-viral area are mostly on sample size re-estimation and dose selection (seamless Phase II/III design)
2. Typically the decision of selection of dose or sample size increase is based on interim analysis results, some times based on a intermediate endpoint (eg Week 16 % BLQ) instead of final endpoint (e.g. Week 48 % BLQ). The interim tend to be early (20% of subjects)
3. Dynamic allocations (Pocock) has been used for some applications. Response adaptive design not so far.

Adaptive vs. Conventional

1. Conventionally, we learn from Phase II about the dose and effect size etc., then move to Phase 3 to confirm the hypothesis. Usually by powering two studies at two-sided 0.05 level
2. Adaptive design, learn within trial and integrate learning phase with confirmatory phase to
   - Increase power
   - Shorten the development time

Pros and Cons

1. Adaptive design has well defined procedure to integrate evidence from early phase with late phase
2. Conventional Phase 2 learning is less structured and therefore may lack a well- and pre-defined procedure for integrating the evidence
   - Not insurmountable
3. But adaptive design may require decision making from DMC with no input from FDA and sponsor
   - A blackbox approach or "give new baby to outsiders"
   - but may not be optimal, or may "not even be suboptimal"
     - May be ok for sample size adjustment
     - More of a problem for dose selection
4. Conventional design allow both sponsor and FDA to fully evaluate the phase 2 and other study results before proceeding to phase 3. Which is a more informative decision and could be better
5. Adaptive design has the potential of shortening the development cycle

Collective Evidence in Conventional Designs

- Approval usually is based on two successful confirmatory trials, with Phase II supportive information. So the overall evidence is usually > 2 confirmatory trials
  - Seamless Phase II/III designs are usually powered for evidence equivalent to two confirmatory trials because of lack of standalone Phase II trials
- Why Phase II evidence matters?
  - Phase 2 evidence level varies greatly from application to application
  - Applications with substantial supportive information should be viewed more favorably. Weak or negative Phase 2 studies should also be taken into consideration.
- Collective evidence requires applicant to submit all relevant studies, no pick and choose
- Collective evidence encourage good early phase and more expansive late phase development, particular better dose finding as the information generated can be integrated for evidence

Dose Finding in Early and Late Phase

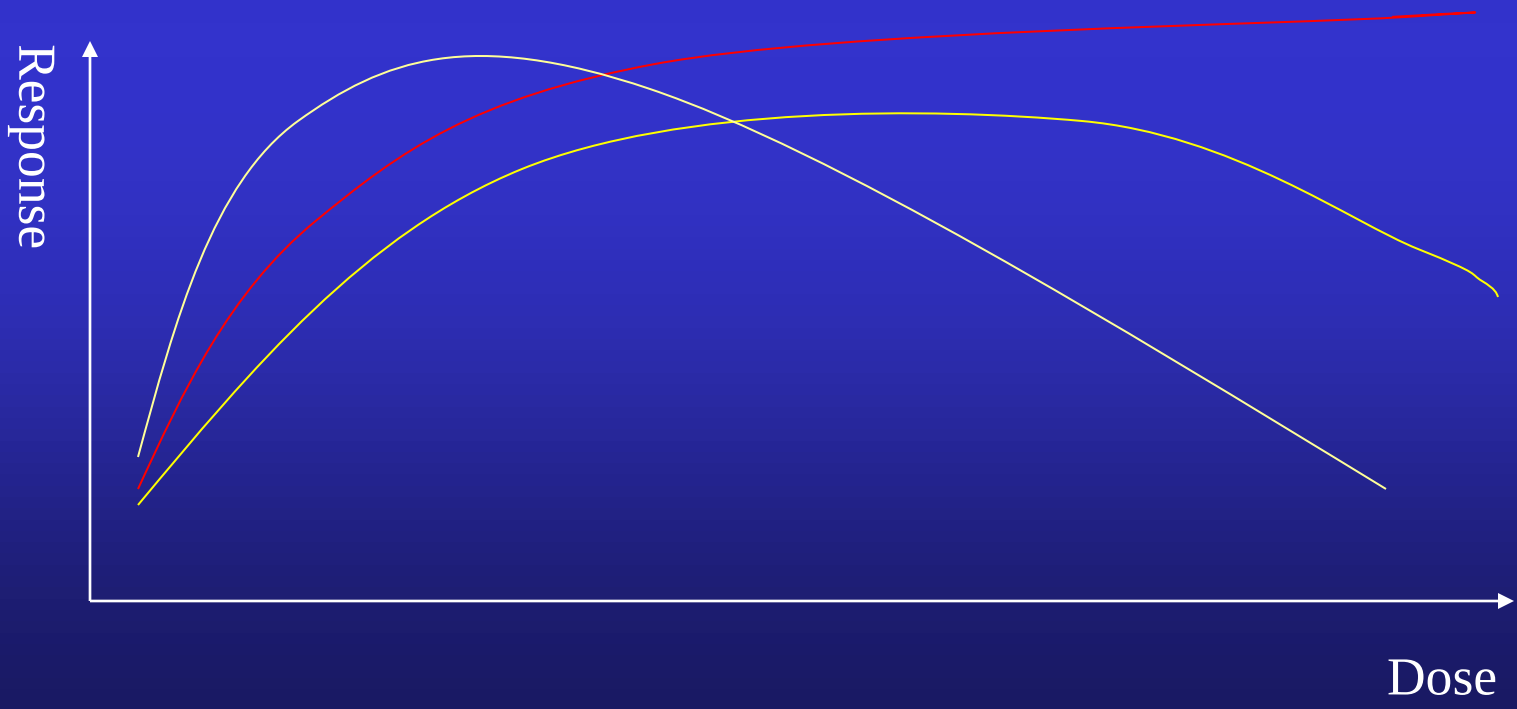
- Goal is not just to find a working dose, goal should be to find a dose with best safety and efficacy balance
- Dose response curve is better estimated by spreading the doses across range, may use (Bayesian) adaptive design approaches to narrow or expand the search
- It could be valuable to carry dose finding to Phase III, study a few doses based on the plausible dose range found in Phase II trials instead choose a single dose prematurely
  - Maraviroc 1027 and 1028 have QD, BID vs placebo
  - More efficient than studying each dose separately because of shared control
  - Allows more comparisons to be made (QD vs pbo, BID vs pbo, QD vs BID)
- Consistency in dose response can be regarded as evidence

Multiple Doses: Can it be more efficient

- Current approaches in handling Phase III multiple doses are either Bonnforoni or step-down.
- Bonnforoni tends not to take good neighboring results as supportive
- Step-down procedure assumes a monotone dose response curve, may not be correct.
  - Difficult to handle when faced with results contradicting to the assumed trend: either lose the Type I error control admitting bad assumptions being made, or ignore the overwhelming evidence on low dose
  - Monotone dose response relationship is difficult to justify – toxicity and AE can affect adherence on high dose more resulting loss of efficacy

Multiple Doses: Can it be more efficient

- However, umbrella shaped dose response curve could be reasonable and can be used to lessen penalty on multiple comparisons

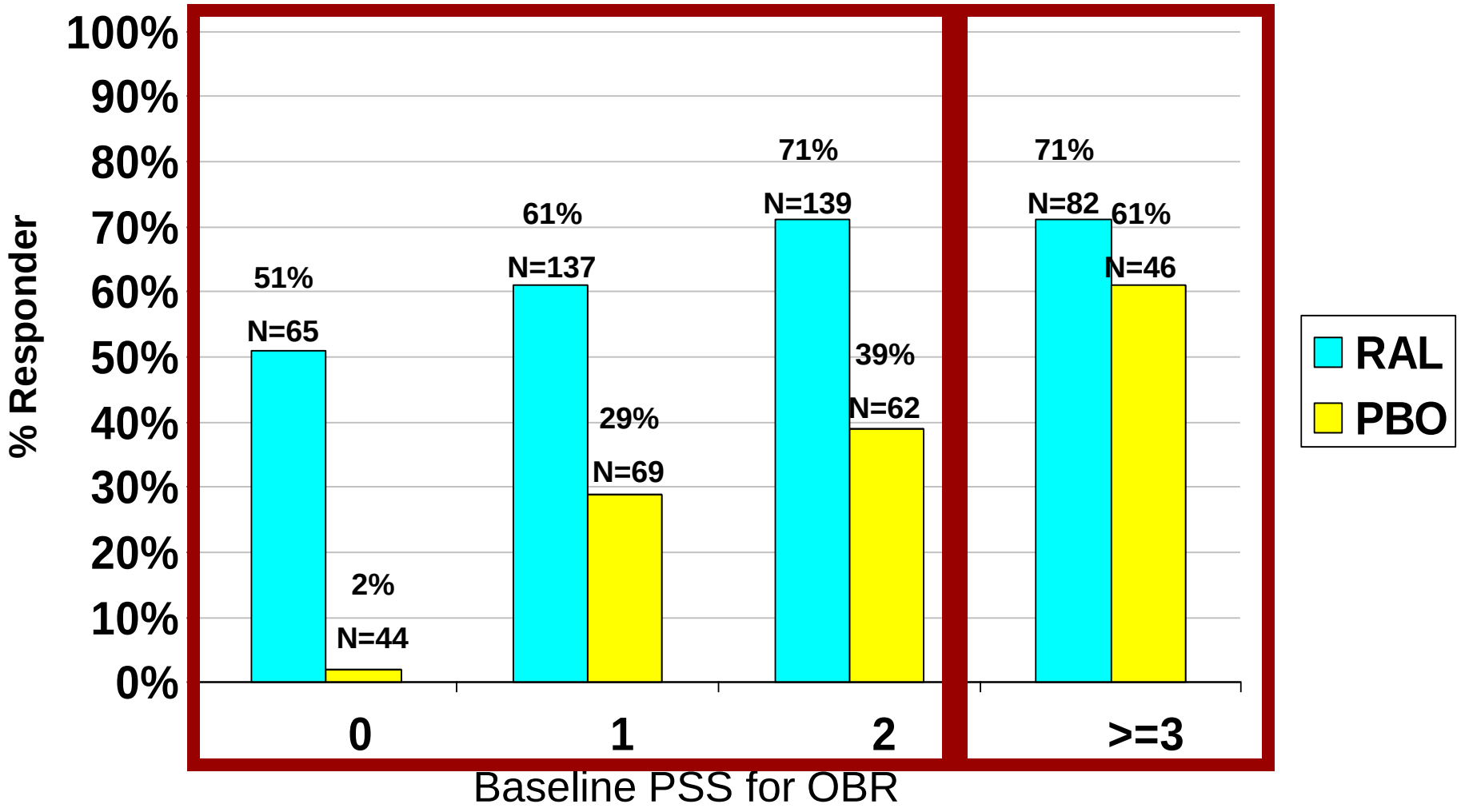


LJ's PIP

- Conditional power may not be a good tool
  - Especially in early stage observed effects are not reliable
  - Effects may also vary over time
- Predicted Interval Plot (PIP)
  - Combine observed data with assumptions on the future for prediction
- Reasonable tool for futility
  - Given the data at interim, assume the future is somewhat brighter than the past, still do not have a decent chance for success
- Can be used to see the need for sample size or duration adjustment
  - Moderate chance of success can be improved by sample size or duration increase
  - How to allocate alpha?

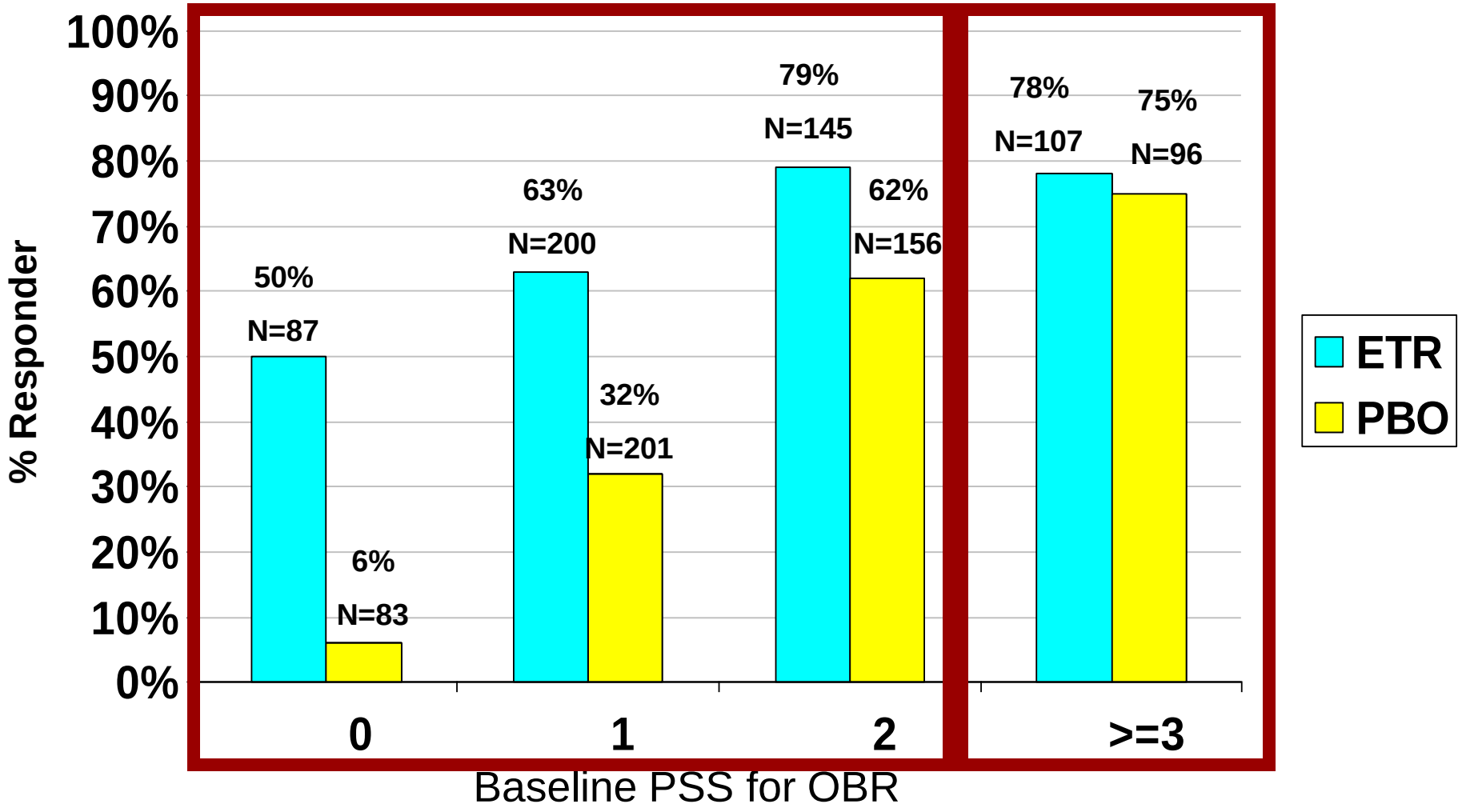
Backup Slides

Baseline PSS vs. HIV RNA Response (%<50)
Pooled BENCHMRK (RAL) Trials
ITT population – Week 48



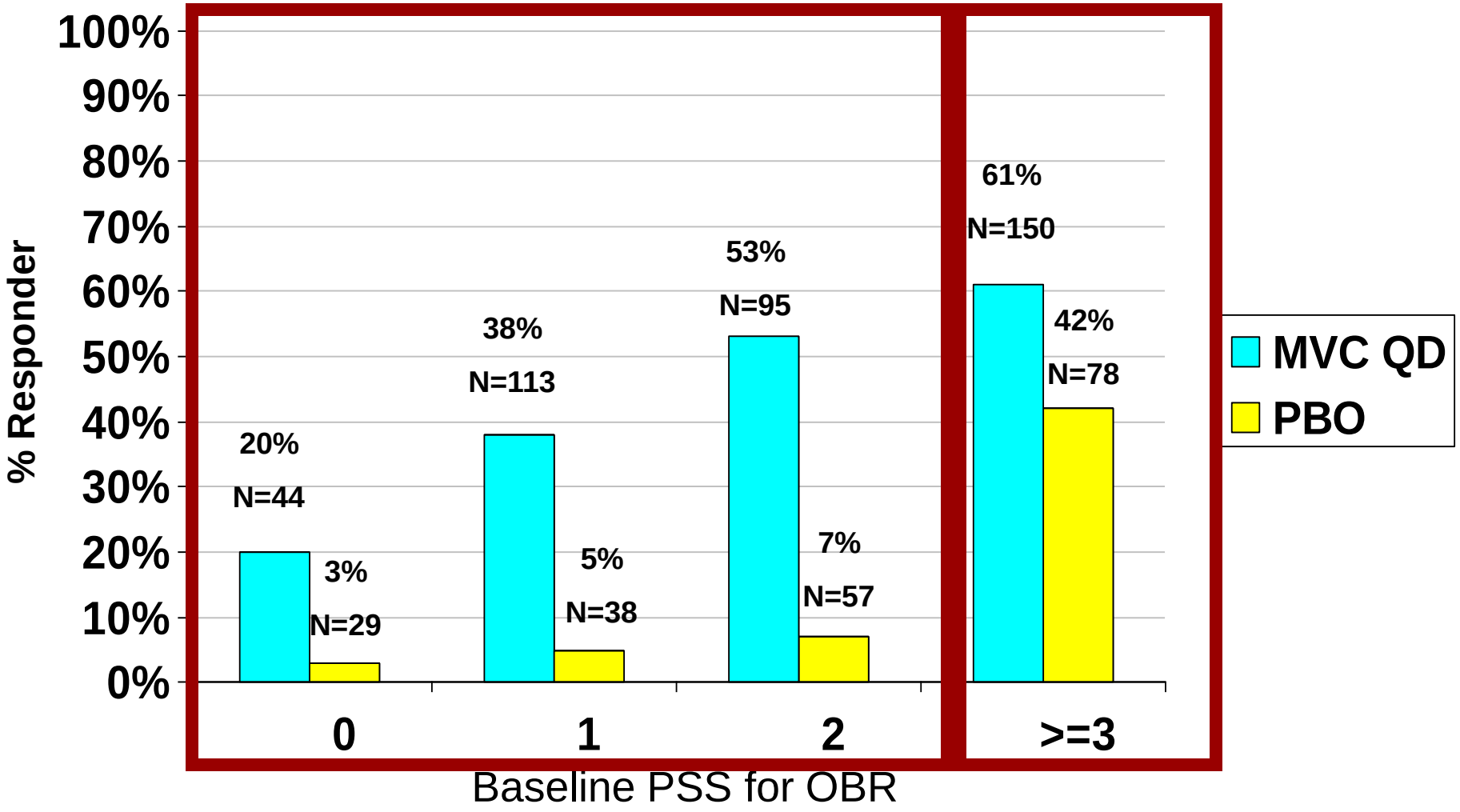
Cooper et al, NEJM 2008;359;355-65

Baseline PSS vs Week 48 Response (%<50)
Pooled DUET (ETR) 1 and 2 Trials*



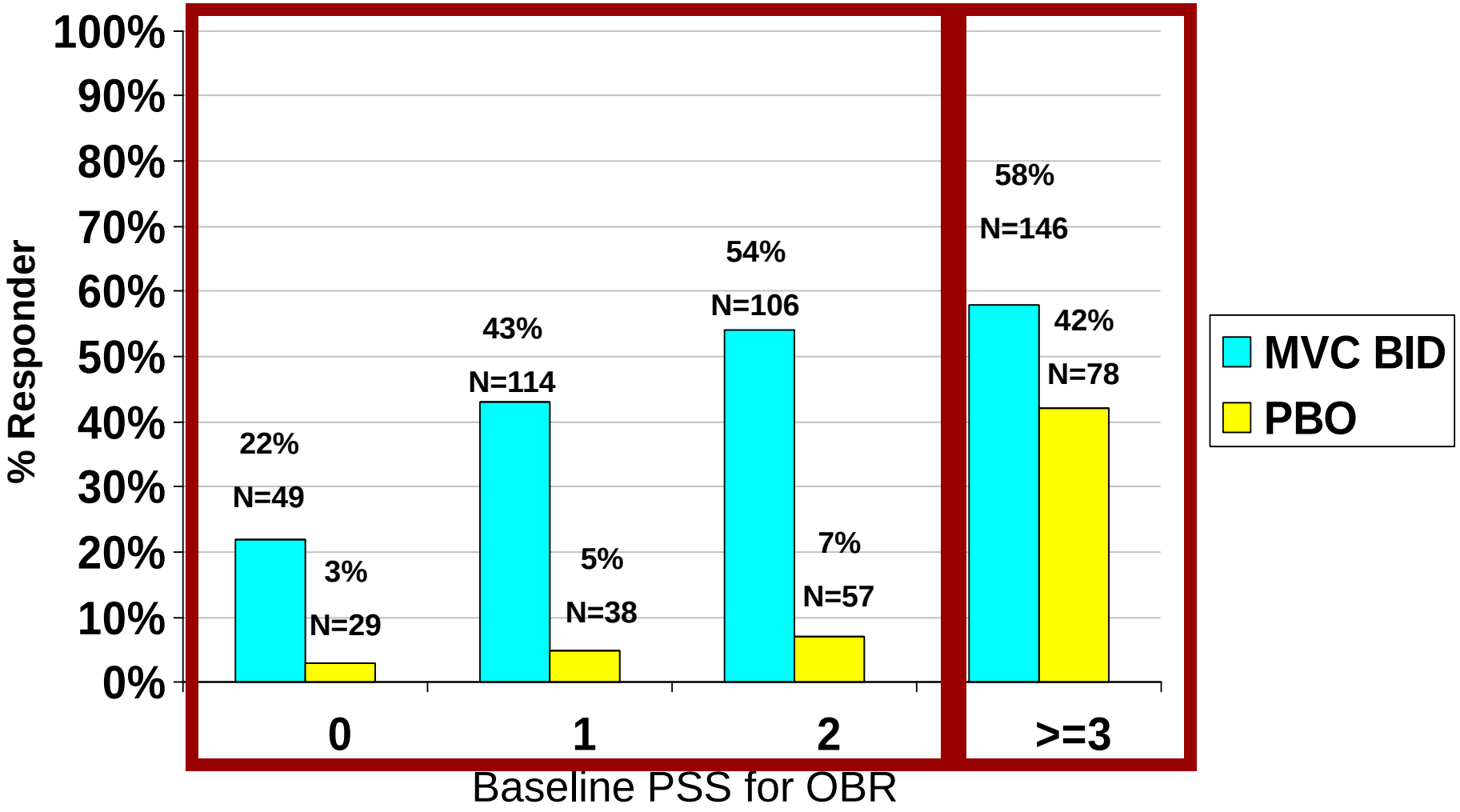
ETR Module 2.7.3 Summary of Clinical Efficacy. Figure 50.

Baseline PSS vs. Week 48 Response (%<50)
Pooled MVC QD 1027 and 1028 Trials*



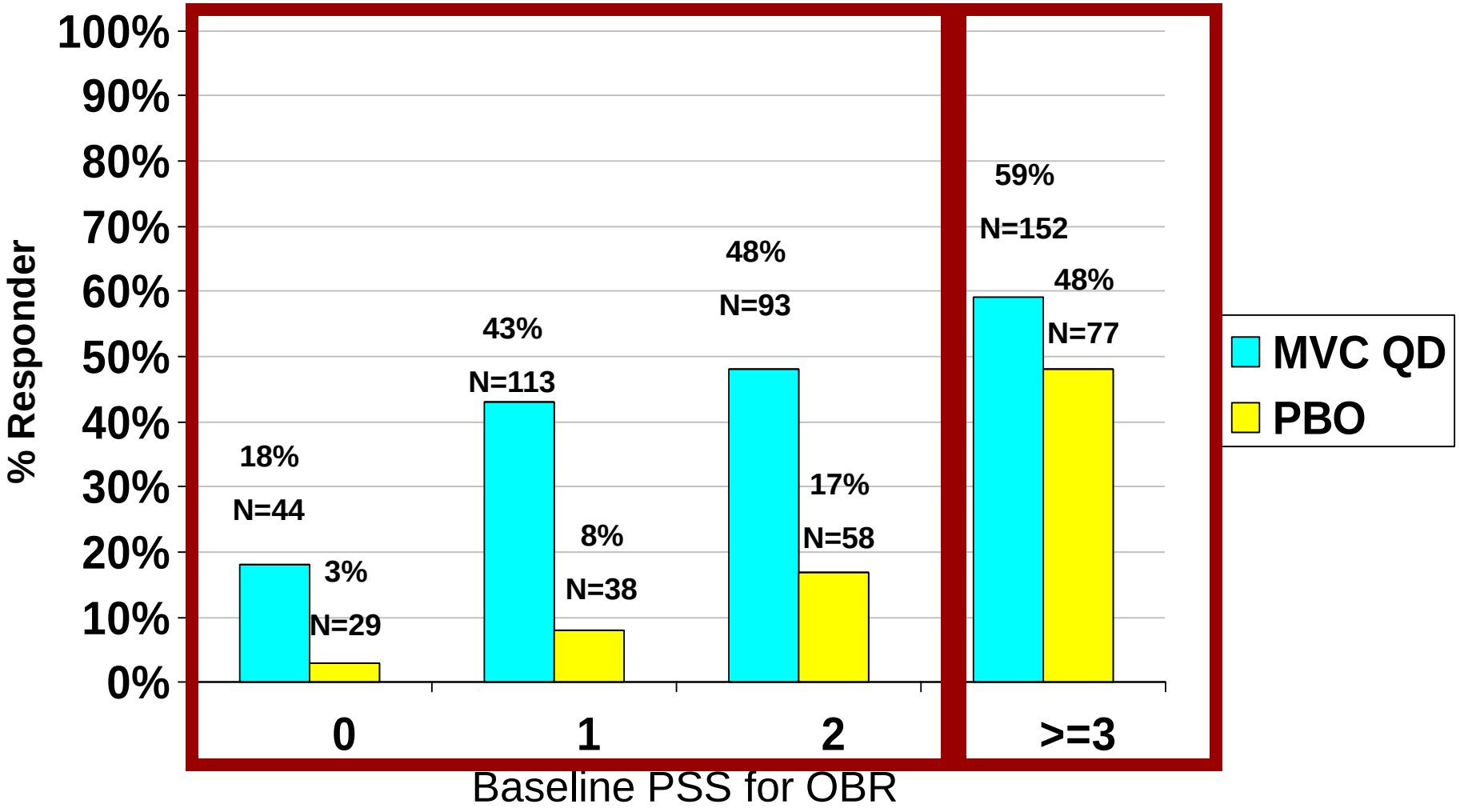
MVC Module 5.3.5.3 Integrated Summary of Efficacy – Week 48. Table 54.

Baseline PSS vs. Week 48 Response (%<50)
Pooled MVC BID 1027 and 1028 Trials*



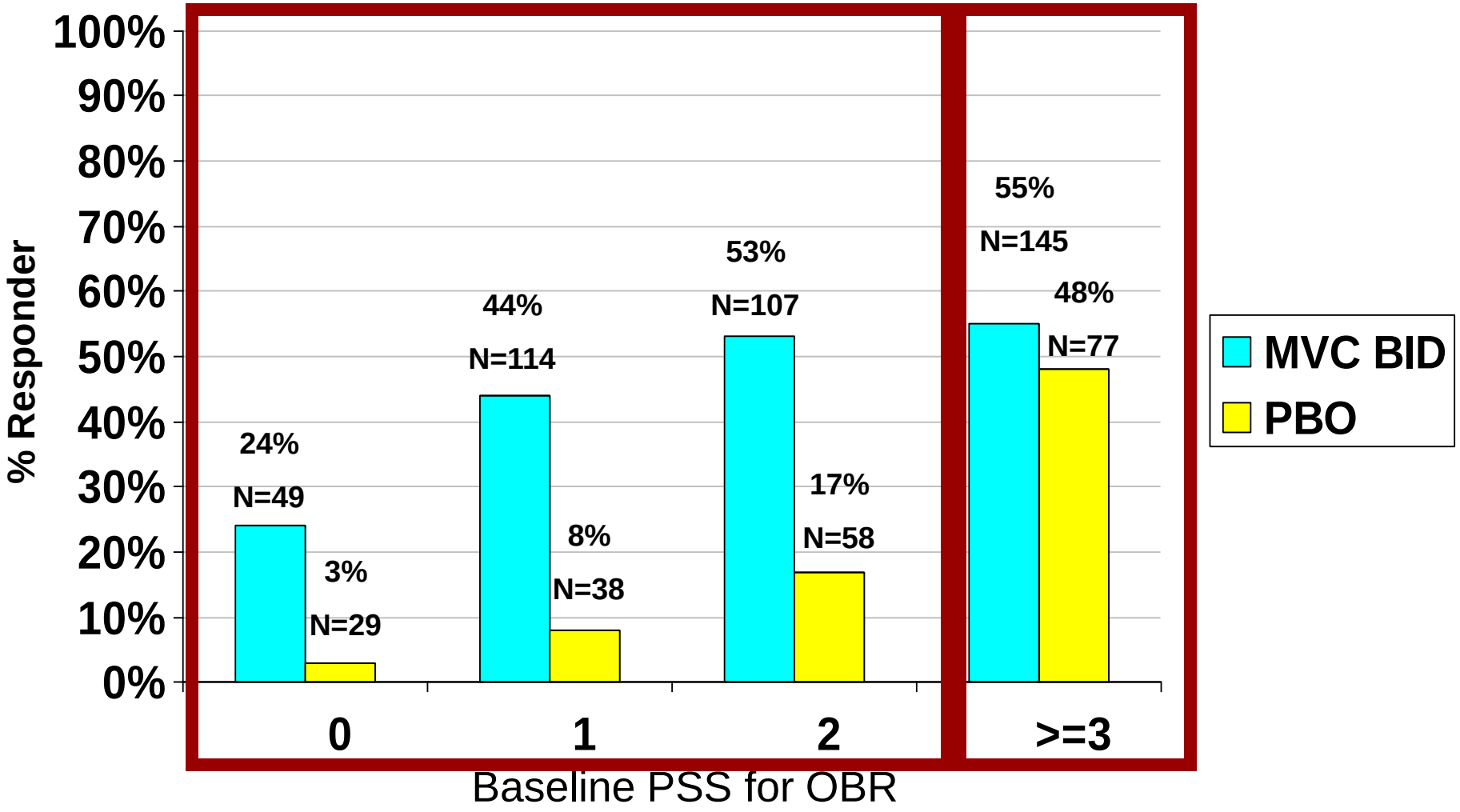
MVC Module 5.3.5.3 Integrated Summary of Efficacy – Week 48. Table 54.

Baseline PSS vs. Week 24 Response (%<50)
Pooled MVC QD 1027 and 1028 Trials*



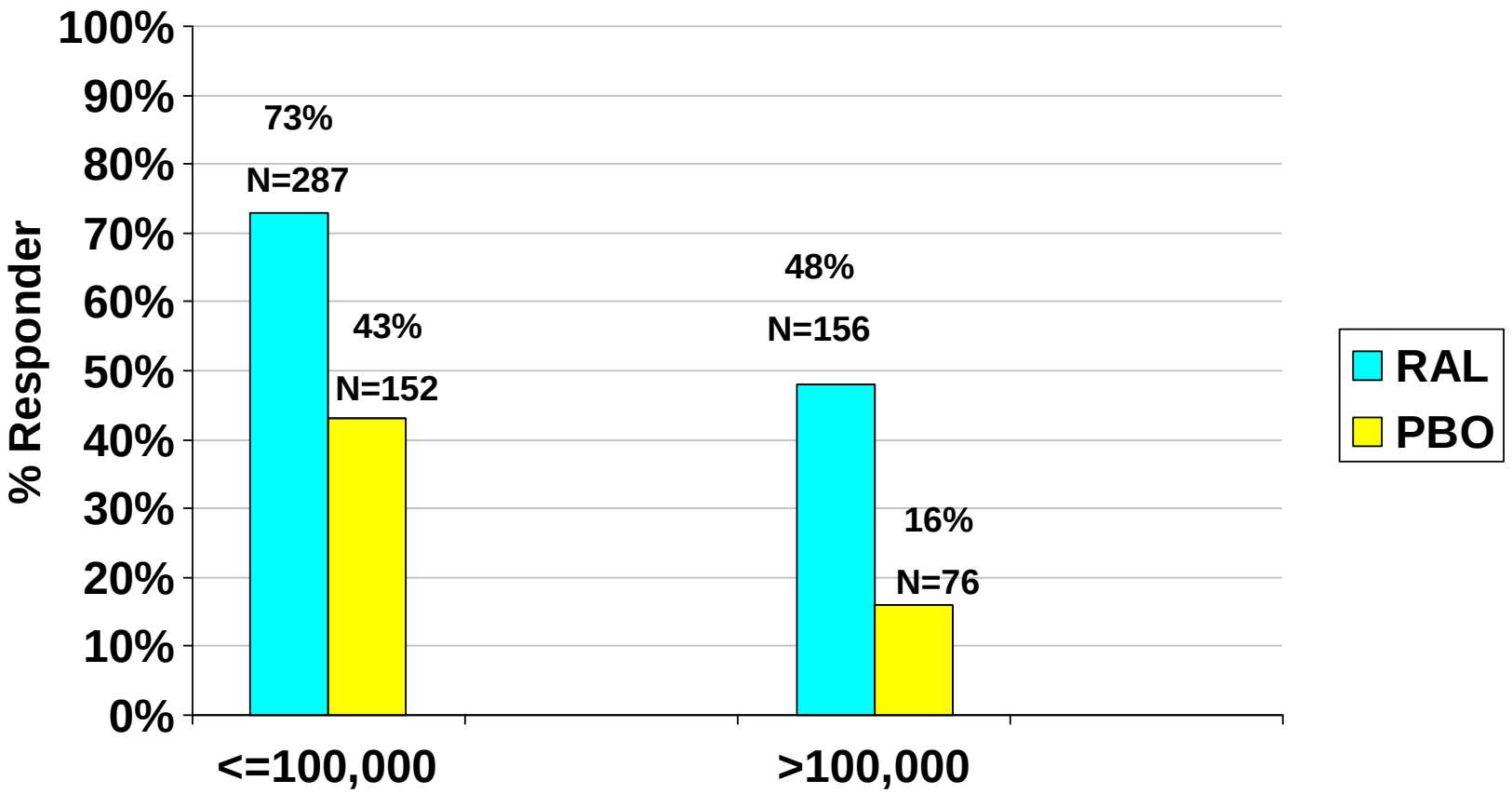
MVC NDA 22,128, AC Briefing Doc. Apr. 24, 2007 from Pfizer Table 89.

Baseline PSS vs. Week 24 Response (%<50)
Pooled MVC BID 1027 and 1028 Trials*



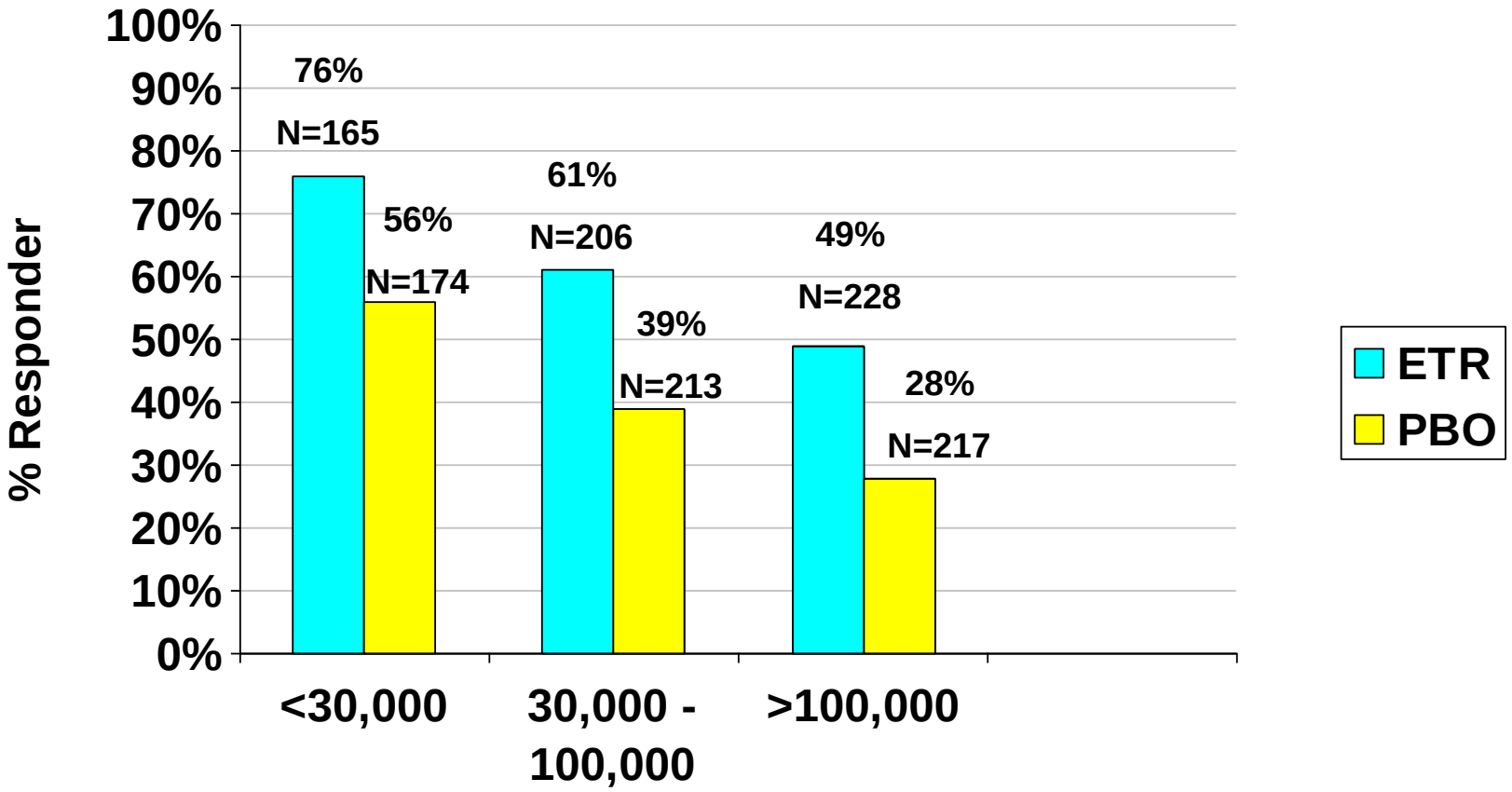
MVC NDA 22,128, AC Briefing Doc. Apr. 24, 2007 from Pfizer Table 89.

Baseline VL vs. HIV RNA Response (%<50)
Pooled BENCHMRK (RAL) Trials
ITT population – Week 48



Baseline HIV-1 RNA Viral Load

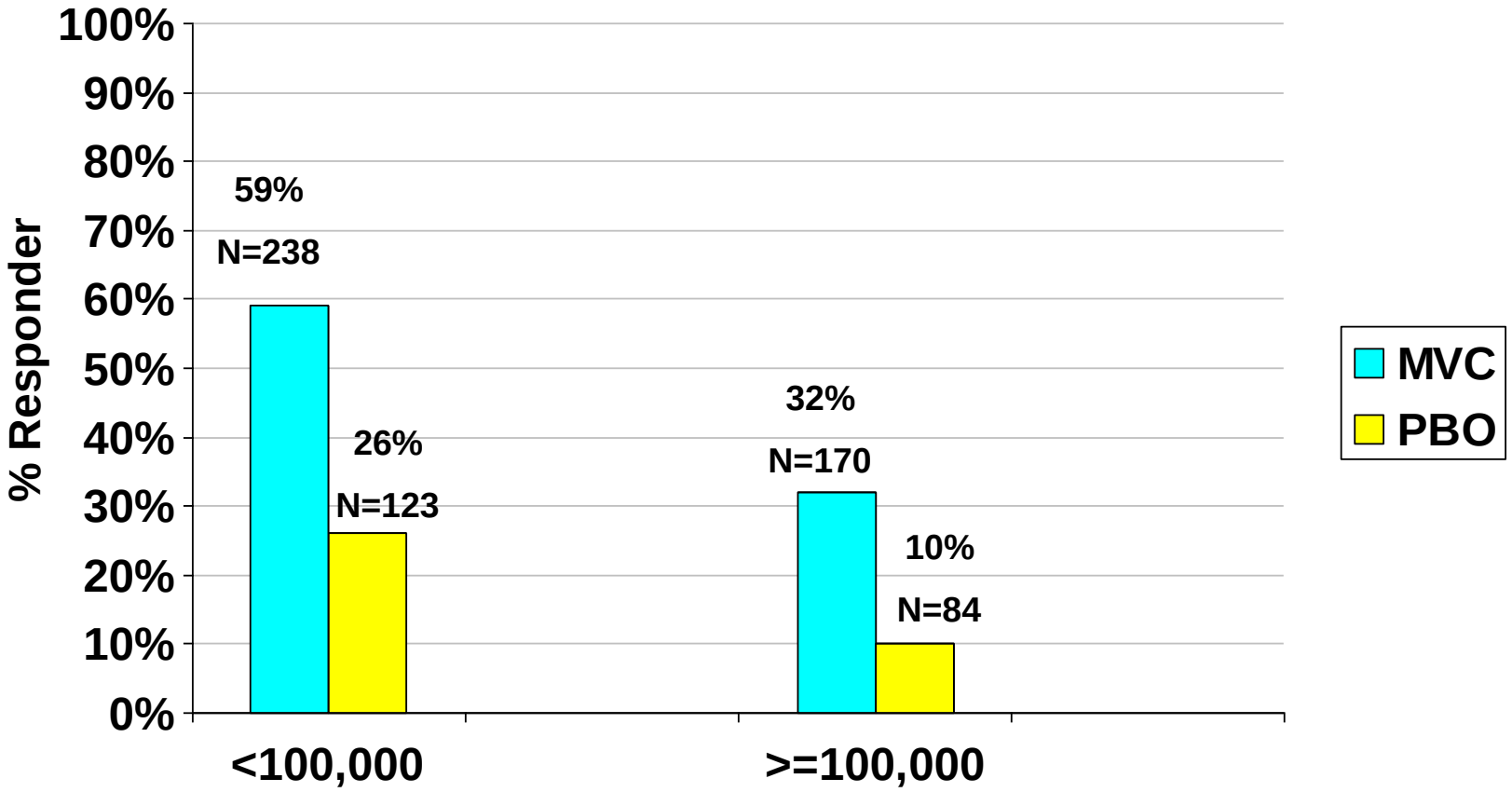
Baseline PSS vs Week 48 Response (%<50) Pooled DUET (ETR) 1 and 2 Trials*



Baseline HIV-1 RNA Viral Load

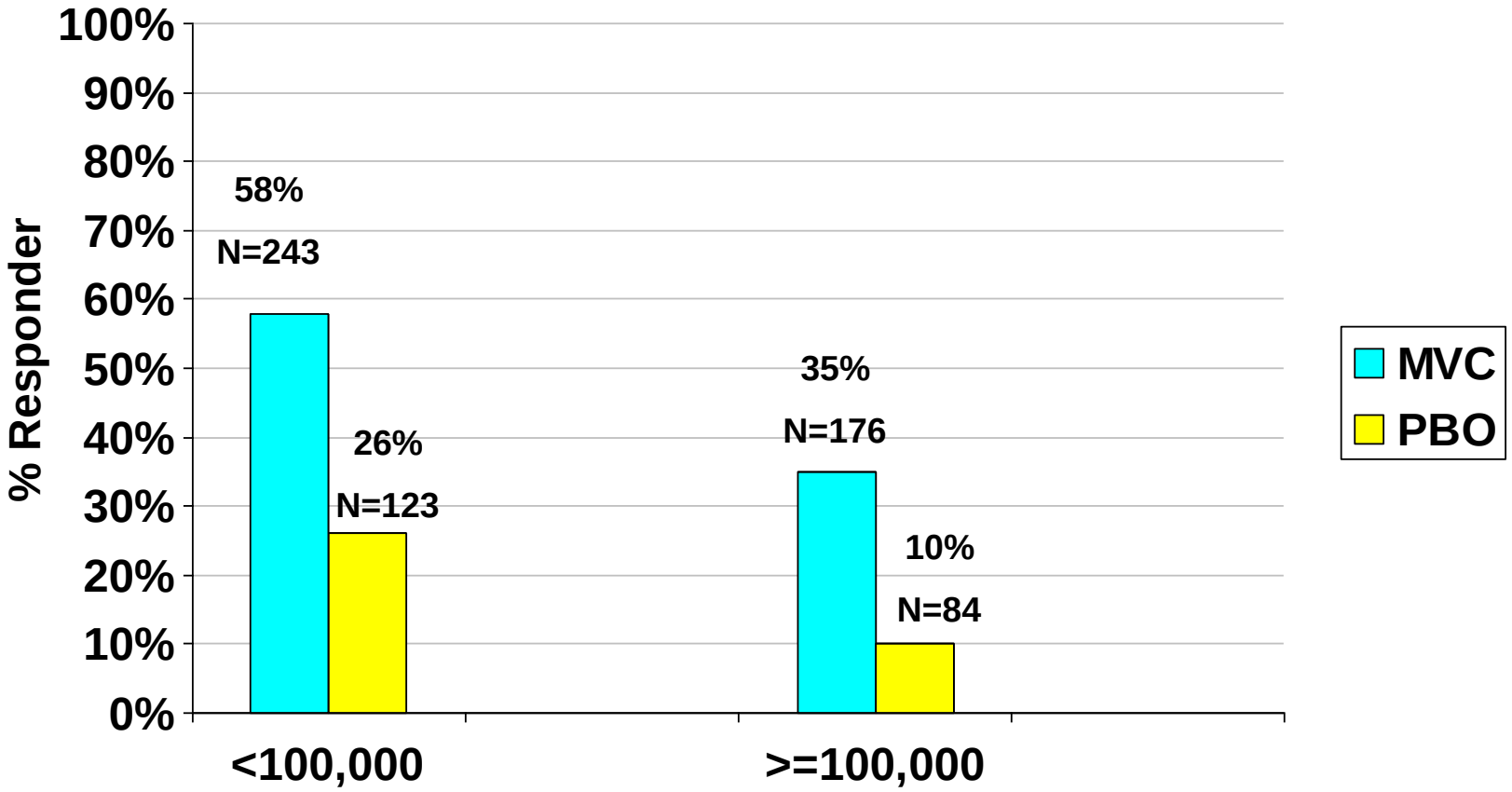
2.7.3 SCE, Figure NDA 22,187

Baseline PSS vs. Week 48 Response (%<50)
Pooled MVC QD 1027 and 1028 Trials*



Baseline HIV-1 RNA Viral Load

Baseline PSS vs. Week 48 Response (%<50) Pooled MVC BID 1027 and 1028 Trials*



Baseline HIV-1 RNA Viral Load