Experience with the Validation of Surrogate Endpoints in HIV

EMEA / CHMP Biomarkers Workshop

London, December 2005

Michael D. Hughes, Ph.D.
Department of Biostatistics
Harvard School of Public Health
mhughes@hsph.harvard.edu
Experience with the Validation of Surrogate Endpoints in HIV

• Historical perspective
  – Major “validation” effort started in 1996
    • New technology for measuring viral load widely available
    • New classes of drug
    • Provided for traditional approval of an antiretroviral drug
      based on effects on a surrogate endpoint

• Highlight issues which may also be relevant to other diseases
What is a Surrogate Endpoint?

“A surrogate endpoint of a clinical trial is a laboratory measurement or physical sign used as a substitute for a clinically meaningful endpoint that measures directly how a patient feels, functions, or survives.

Changes induced by a therapy on a surrogate endpoint are expected to reflect changes in a clinically meaningful endpoint.”

Ref: Temple. In Nimmo, Tucker, eds. 1995
Clinical Endpoints in HIV

- AIDS/death endpoint generally used
  - Definitions varied
  - AIDS events = selected opportunistic infections, cancers, wasting, etc., that characterize immunodeficiency
  - Very heterogeneous
    - Some (e.g. herpes simplex infection) have no clear association with risk of death
    - High relative risk (>5) of death for others (e.g. lymphoma)
Surrogate Endpoints in HIV

- **CD4+ T lymphocyte**
  - A measure of immune status
  - HIV replicates within this cell
  - Lower is worse
  - Typically 800-1000 cells/µL in healthy adults
  - AIDS events typically rare until <200 cells/µL

- **HIV-1 RNA**
  - A measure of viral load
  - Higher is worse
  - Values range from below assay limit of detection (e.g. <400 copies/mL) and up into the millions
  - Often expressed as \( \log_{10} \) copies/mL
Considerations in Defining “HIV-1 RNA” as an Endpoint

- Specimen type (body compartment)
- Choice of assay
- QA procedures for specimen handling/assay
- Measurement time(s) relative to start of treatment
- Statistical analysis
  - Handling of losses to follow-up and deaths prior to measurement
  - Handling of measurements outside assay’s range of quantification
Definitions

• **BROAD ISSUE:** Heterogeneity in definitions impairs ability to evaluate a potential surrogate
  – Ideally, need prospective standardization of definitions of potential surrogate endpoints and clinical endpoints
  – Sometimes achievable retrospectively if patient-level data available
Validating a Surrogate Endpoint

- Need to build up a hierarchy of information about a potential surrogate endpoint:

  1. Should be a prognostic marker
  2. Changes in the potential surrogate after starting a treatment should be prognostic
  3. HARDEST: Effects of treatments on the marker should explain/be associated with effects of treatments on the clinical endpoint

- Biological rationale for a surrogate important
Validating a Surrogate Endpoint

- Need to build up a hierarchy of information about a potential surrogate endpoint:
  
  1. Should be a prognostic marker
  2. Changes in the potential surrogate after starting a treatment should be prognostic
  3. HARDEST: Effects of treatments on the marker should explain/be associated with effects of treatments on the clinical endpoint
Are HIV-1 RNA and CD4 Count Prognostic Markers for AIDS/Death?

- Evaluate in natural history studies
- Pivotal study used data from the Multicenter AIDS Cohort Study (MACS)
  - Started in 1980’s prior to widespread treatment
  - CD4 counts routinely measured
  - Stored specimens available to measure HIV-1 RNA [SIGNIFICANT ISSUE]
- 1604 homosexual men in U.S. cities without AIDS, with specimen
  - Followed for over 9 years
MACS: Percentage Progressing to AIDS/Death in 3 Years by HIV-1 RNA and CD4 Count

Validating a Surrogate Endpoint

• Need to build up a hierarchy of information about a potential surrogate endpoint:

(1) Should be a prognostic marker

(2) Changes in the potential surrogate after starting a treatment should be prognostic

(3) HARDEST: Effects of treatments on the marker should explain/be associated with effects of treatments on the clinical endpoint
Are Changes in a Potential Surrogate After Starting a Treatment Prognostic?

• Can be evaluated in observational studies of treated subjects and in treatment trials

• Ideally, association should be independent of the treatment used
Evaluation of Surrogacy Needs Major Collaborative Effort

• HIV Surrogate Marker Collaborative Group
  – Major collaboration of government-sponsored trials groups and pharmaceutical companies

• Forum for the evaluation of HIV-1 RNA and CD4 count as surrogate endpoints in HIV clinical trials for progression to AIDS/death

• Led to several cross-study analyses

• Contributed to FDA and EMEA policy discussions
HSMCG Meta-Analysis

• All 16 trials with HIV-1 RNA measured and involving one class of drugs (NRTIs)
  – Markers: Change in HIV-1 RNA and CD4 cell count over 24 weeks
  – Clinical endpoint: Progression to AIDS/death over 2 years

• 13,045 patients
  – 3369 (26%) developed AIDS or died
  – 3,146 patients with HIV-1 RNA measurements (often done retrospectively using stored specimens)

[ref: HSMCG. Aids Research and Human Retroviruses, 2000]
HSMCG: Prognostic Value of Changes in HIV-1 RNA and CD4 Count

(a) Percentage reduction in hazard (log scale) vs. Change in log10 HIV-1 RNA

(b) Percentage reduction in hazard (log scale) vs. Percentage change in CD4 (log scale)
HSMCG: Prognostic Value of Changes in HIV-1 RNA and CD4 Count

Percentage progressing to AIDS/death

Weeks from start of treatment

- RNA > baseline, CD4 < baseline (n=467)
- RNA < baseline, CD4 < baseline (n=895)
- RNA > baseline, CD4 > baseline (n=307)
- RNA < baseline, CD4 > baseline (n=1477)
Estimated Hazard Ratio for Progression to AIDS/Death by Treatment (for each 1 log10 reduction in HIV-1 RNA during first 24 weeks of treatment)
Estimated Hazard Ratio for Progression to AIDS/Death by Treatment (for each 33% increase in CD4 during the first 24 weeks of treatment)

Adjusted hazard ratio (log scale)

0.79 [0.75,0.83]

Heterogeneity test
Chisq=16.2 on 11 df
p= 0.135
Other Evidence about Marker Changes After Starting Treatment

• Similar associations subsequently found for treatments involving other classes of drugs

• Longer duration of virologic suppression associated with greater CD4 increases and greater reduction in risk of AIDS/death
  – Fits with biological model
Prognostic Early Changes Are Not Sufficient to Validate a Surrogate

- The association vs. causation problem

- Subjects who would have had a better prognosis in the absence of treatment may be more likely to “respond” to treatment
Validating a Surrogate Endpoint

• Need to build up a hierarchy of information about a potential surrogate endpoint:

(1) Should be a prognostic marker

(2) Changes in the potential surrogate after starting a treatment should be prognostic

(3) HARDEST: Effects of treatments on the marker should explain/be associated with effects of treatments on the clinical endpoint
Validating a Surrogate Endpoint: Going Beyond Correlation

• Temple’s definition: “..... Changes induced by a therapy on a surrogate endpoint are expected to reflect changes in a clinically meaningful endpoint.”

• To properly evaluate “changes induced by a therapy”, need a randomized trial, e.g. of the therapy versus placebo
Evaluating A Surrogate in a Meta-Analysis of Randomized Trials

• Evaluate association of the difference between randomized treatments in effect on the clinical endpoint and corresponding difference in effect on the surrogate across multiple randomized comparisons (trials).

• Good to have heterogeneity in treatments

Schematic of a Good Surrogate

Difference in Clinical Endpoint

Difference in Marker
Statistical Model for Meta-Analysis

• For \( i \)th randomized treatment comparison:
  – \( \gamma_i \) = true difference for surrogate
  – \( \theta_i \) = true difference for clinical endpoint

• Model: \( \theta_i = \alpha + \beta \gamma_i + \varepsilon_i \)

• Interpretation:
  – \( \beta = 0 \): marker has no predictive value (i.e. not useful as a surrogate)
  – \( \beta \neq 0 \): want variability of \( \varepsilon \)'s to be small
  – Good rationale for wanting \( \alpha = 0 \) as then no difference in surrogate means no difference in clinical endpoint
Meta-Analysis of HIV RNA as a Surrogate
Meta-Analysis of HIV RNA as a Surrogate

(1) Few qualitatively discordant treatment comparisons
Meta-Analysis of HIV RNA as a Surrogate

(2) Quantitative discordance
Meta-Analysis of HIV RNA as a Surrogate

Log Hazard Ratio of AIDS/death

Difference in HIV RNA (log copies/mL)

(3) Estimates: $\beta = 0.28 \; (-0.16, 0.70)$

$\alpha = -0.12 \; (-0.34, 0.08)$
HIV-1 RNA as a Surrogate Endpoint

- Different metrics for measuring suppression of HIV-1 RNA
  - For area under curve minus baseline (AUCMB), trend, $\beta$, marginally significant

- Clinical interpretation: sustained suppression more important as a surrogate than simple change
Meta-Analysis of CD4 Count as a Surrogate

Estimates: \( \beta = -4.1 \ ( -7.1, -1.6) \)
\( \alpha = 0.04 \ ( -0.16, 0.29) \)
HIV-1 RNA and CD4 as Joint Surrogates

- Include both HIV-1 RNA (AUCMB) and CD4 in meta-analysis regression model:
  - $\beta_{RNA} = 0.07$ (-0.49, 0.59)
  - $\beta_{CD4} = -3.9$ (-7.7, -0.5)
  - $\alpha = 0.04$ (-0.20, 0.31)

- Given effects on CD4, little additional predictive value of HIV-1 RNA
  - Biologically, CD4 is closer than HIV-1 RNA on the pathway to AIDS/death
Meta-Analysis Extended to Include Other Classes of Drugs

- Used publicly available summary data
  - Difficulties accessing individual patient data
- Marker changes evaluated at 16 weeks
- Used all available follow-up for AIDS/death
  - Less standardization

Meta-Analysis Including NRTI, NNRTI and PI Drug Classes

- Both associations significant

FDA Guidance for Approval of Antiretroviral Drugs

• Accelerated and traditional approval based primarily on effects on HIV-1 RNA
  – 24 weeks for accelerated
  – 48 weeks for traditional

• Supporting data on effects on CD4 count and clinical endpoints, particularly for traditional approval
Why HIV-1 RNA as the Basis for Evaluating New Drugs?

• Highly active antiretroviral therapy (HAART) became available at about same time as technology for measuring HIV-1 RNA

• HAART reduced HIV-1 RNA on average by $2 \log_{10}$ copies/mL or more vs. previous therapies
  – Effect rapid – within a few weeks

• HIV-1 RNA assays very sensitive ……
Short-Term Within-Subject Variability in HIV-1 RNA Without Treatment Changes
Why HIV-1 RNA as the Basis for Evaluating New Drugs? - Practicalities

- Effect of HAART on HIV-1 RNA can be observed within individual patients
  - Conversely: loss of effect indicating treatment failure
- Patients often switch treatments when lack or loss of effect on HIV-1 RNA observed
- Essentially impossible to conduct RCTs of HAART regimens using clinical endpoints
  - Also very difficult with CD4 as endpoint
Why HIV-1 RNA as the Basis for Evaluating New Drugs? - Surrogacy

• HIV-1 RNA is a measure of pathogen level causing the disease

• Clear that treatment-mediated changes in HIV-1 RNA while on HAART sufficiently large that clinical benefit is obtained

• Advent of HAART in 1996 associated with subsequent dramatic effect on mortality observed in disease surveillance ……
Incidence, prevalence and deaths among persons with AIDS, 1985 - 2001, United States

- **AIDS**
- **Deaths**
- **Prevalence**

Data adjusted for reporting delays and for estimated proportional redistribution of cases reported without a risk; data reported through June 2002
Why HIV-1 RNA as the Basis for Evaluating New Drugs? - Surrogacy

• Little evidence of qualitative discordance in association of difference between treatments in effects on markers and corresponding difference in effects on AIDS/death

• RCTs show that prophylaxes for opportunistic infections can be withdrawn safely following virologic suppression and increased CD4 count on HAART
Surrogate Endpoints in HIV: Some Limitations

• Use of a surrogate necessarily involves an extrapolation of past experience
  – Now evaluating new classes of antiretroviral drugs based on HIV-1 RNA
  – Not applicable to other types of treatment, e.g. immune-based therapies

• Appreciable quantitative discordance, particularly for HIV-1 RNA
  – May be most relevant for patients for whom treatment options have limited effect
  – Need for treatment management trials with clinical endpoints
Surrogate Endpoints in HIV: Some Limitations

• Data almost exclusively for HIV-1 subtype B
  – How applicable to subtypes most prevalent in Africa etc.?

• Data largely from patients with later stage infection
  – What is best surrogate for evaluating secondary effects of vaccines among subjects who do become infected?
Evaluating Surrogate Endpoints in HIV: What Helped?

• Multiple potent treatments

• Sensitive assay for HIV-1 RNA (relative to treatment effect)

• Very extensive database
  – Many RCTs with large number of clinical endpoints
  – Ability to evaluate HIV-1 RNA retrospectively in a large number of patients using stored specimens

• Reasonable biological model

• Strong collaborative effort
Challenges for Validating Surrogates

• Need standardized definitions of potential surrogates and clinical endpoint

• Need systematic evaluation of potential surrogates
  – Large databases (obs. studies and trials), particularly if treatments not very potent
  – Collaborative effort important

• Need new methodology for linking information from shorter-term studies when clinical endpoints are distant in time
  – Particularly if patients discontinue therapy prior to clinical endpoint at a high rate