

The Crisis in AIDS Research *by Mark Harrington*

An overview of AIDS research, written in 1993, with suggestions for changes in fundamental approaches to how the virus, treatments, and research goals are viewed.

THE CRISIS IN CLINICAL AIDS RESEARCH

by Mark Harrington

1 December 1993

*

Telling people with AIDS the truth isn't taking away their hope -- it's giving them the information they need to make informed treatment decisions.

-- Sandra Johnson CCG/Women's Health Committee

*

In Memoriam. This report is dedicated to the memory of those who served on the ACTG Community Constituency Group (CCG) and who are no longer alive: Ortez Alderson, Jesse Dobson, Dick James, Chuck Mayer, Bob Pearson, Jose Perez, Maritza Ramos, and Andy Zysman.

*

Introduction. In November 1989, AIDS activists attended our first AIDS Clinical Trials Group (ACTG) meeting. We were not given a warm welcome. After many months of struggle, activists and people with HIV were integrated into the ACTG system and given voting representation on each committee, and many ACTG sessions were opened up to community observers.

Four years after those events, however, many of the initial criticisms of ACTG research remain relevant. Moreover, new problems have emerged which demand changes in the design, conduct and analysis of clinical trials for AIDS and HIV-related conditions. In addition, hard-won victories by activists (e.g., parallel track, active-controlled studies, accelerated approval) have produced a new set of issues of their own which must be addressed. Finally, results of recent, relatively large-scale clinical trials have been disappointing; despite substantial aggregate improvements in quality and length of life since the mid-1980s, the hoped-for long-term benefits of early or combination intervention with antiretroviral therapies have not materialized.

Those designing the latest batch of clinical trials have tended to ignore these setbacks, proceeding as though nothing was wrong with the assumptions, some of which are now undermined by the data, underlying these studies. This report is an effort to critically assess the current impasse.

A Few Words of Praise. Many of the interactions between activists and researchers within the ACTG and elsewhere have been positive, productive, and indeed historic. I myself was lucky enough to serve on the ACTG Opportunistic Infections (OI) Core Committee in 1990-91, when for the first time (and following concerted pressure from activists and from Congress) appropriate resources were being allocated to the OI prophylaxis and treatment effort. In my experience, the OI Committee did not suffer internally from the ego and turf problems which characterize some other ACTG committees. Both Core discussions and those in Pathogen Study Groups (PSGs) focused on the problems at hand, and debate was full and frank.

The recent ACTG Scientific Review panel commended the OI committee's "young, energetic, high-quality investigators", noting that "the OI Committee has the

broadest scope and charge of the ACTG committees... has focused on the most prominent clinical problems in AIDS... conducted well-conceived, comparative studies that had power and led to clear conclusions [and]... contributed significantly to standards of care and a better quality of life for AIDS patients... The program is moving in the right direction -- of co-enrollment, collaborations with CPCRA [and others], and rollover studies... prophylaxis of multiple pathogens and... polypharmacy." The OI Committee, once given appropriate support from NIAID, has contributed greatly to the standard of care for AIDS and HIV disease. As they shift focus from an emphasis on treatment to one on prophylaxis, they too will face many of the problems related to large sample size, long duration, losses to follow-up, subset analysis, etc., now afflicting the primary infection effort.

Other ACTG committees too have made progress since 1989. The Oncology Committee, with new support from the NCI AIDS Lymphoma Network, and brilliantly advocated for by the late Andy Zysman, has created a comprehensive research agenda. Stronger support from NIAID (including an oncologist medical officer) is needed.

The Neurology Committee still receives inadequate support from NIAID and from many site principal investigators (one Neurology chair resigned because his site PI failed to support his work), but has a new arrangement with the National Institute of Neurological Disorders and Stroke (NINDS) to collaborate in a Neurology AIDS Research Consortium (NARC). Hopefully, this may lead to the sort of synergistic collaboration which the NCI/Oncology Committee link now provides.

The Immunology Committee has spread its wings from its initial role as a laboratory-based, resource committee to once conducting increasingly important but still preliminary studies of immune-based therapies for HIV infection. Again, the pointed activism of CCG Immunology representative Jesse Dobson was critical in mobilizing this talented group. Like Oncology and Neurology, the Immunology Committee still receives inadequate support from the Division of AIDS (though the tireless work of medical officer Jonathan Kagan deserves high praise).

Rearranging the Deck Chairs. The new ACTG restructuring proposal ratifies the virtually complete separation of adult and pediatric components. In its new HIV/AIDS Research Agenda, NIAID separates out pediatrics research, apart from pathogenesis, epidemiology, clinical trials, and vaccine research, as though children were a separate species.

Among those who will suffer from the separation, which NIAID has done nothing to discourage, are 1) mothers whose HIV-infected children will be treated at pediatric ACTUs where there are no research facilities for the adults in the family; 2) adolescents with HIV who will fall between the logistical and eligibility cracks in protocols and programs which focus either on adults (many drug companies don't like to have patients under 18 in their studies) or children (some pediatric studies cut off enrollment at age 12); 3) HIV infected children themselves. This is for several reasons: The Pediatrics Committee's agenda privileges interruption of vertical transmission, and antiretroviral therapy, over OI prophylaxis and treatment. There are too few children to fully enroll many efficacy studies at statistically necessary levels. The Pediatrics Committee is studying many means of potentially interrupting vertical transmission (e.g., AZT in ACTG 076, envelope-based vaccines in ACTG 233-235), but too few dealing with the complications of HIV in children post-infection. They appear to feel that pilot studies of envelope-based immunogens offer more clinical promise than OI prophylaxis or treatment studies. As the Pediatrics Committee splits off to form its own wasteful, duplicative, redundant structure, the remainder of the ACTG is now dubbed "Adult". Yet only these supposedly

"adult" research committees will be able to obtain sufficient sample size to obtain definitive efficacy information for treating HIV-infected persons of all ages.

Meanwhile, the "adult" system lacks rigor and vision. Protocols are not costed out properly. The committees have their own research agendas, but the group as a whole does not. No one at DAIDS knows where all the dollars go. Efforts to reduce collection of unnecessary data have been token at best. In the meantime, the ACTG cannot follow most of its study participants to the end of the trials. This cripples large-scale, long-term studies, like that of AZT in persons entering with CD4<500/mm³, ACTG 019, of whom over three-quarters had left before the study closed out this fall. The Executive Committee has, according to one member, already approved trials for 1994 which will take up the entire year's resources. Thus, no new ideas can be tested in 1994, unless some currently planned trials are dropped. Drop them, I say.

The new Executive Committee will give more power to the site principal investigators, from whom the majority of its members will be drawn. The EC will remain ineffective, however, unless new systems for budgeting trials up front and tracking costs throughout the system are implemented. No EC composed of site PIs can administer the far-flung ACTG apparatus without responsible, committed staffing by DAIDS. Hopefully the new Division Director will recognize this.

The Crisis in Clinical Trials. Still, the current dilemma of treatment research in AIDS extends far beyond the ACTG system.

These are the components of the dilemma: 1) inadequate drugs; 2) inadequate markers; 3) inadequate clinical trials. After the Berlin conference, SDAC honcho Steve Lagakos commented to one community member that "those activists wouldn't be so mad [about the much-hyped "subset trend analysis" from ACTG 155] if the drugs were better." Well that's right, Steve! -- if the drugs were better, then the trials wouldn't need to be bigger, or better designed, or analyzed more honestly -- in fact, if the drugs were good enough, we might not need answers from randomized studies at all, as in the case of ganciclovir. But the drugs aren't better -- and that's why we turn to statisticians in the hope that they will help design studies competently, and analyze them honestly, keeping in mind that the primary goal is the development of information useful to patients and their providers.

The fact is, that of the three inadequacies listed above, clinical trials designers have the most control over the third. They cannot invent good drugs -- that is the job of basic researchers; they cannot invent better markers -- though they can help to develop them; they can, however, design trials adequate to answer clinically relevant questions, and they can do so without resorting to exaggeration, hype, or well-meaning distortion in the face of disappointing overall results. People with HIV obviously want, and need, better drugs. But until those come along, it is imperative to use the drugs we have now better, to provide clearer information to people with HIV and their providers, so that we can make the best use of what is out there.

Three Case Studies. It might be useful to approach the issue by reviewing three episodes which exemplify troubling aspects of the design, conduct, analysis and interpretation of clinical trials for HIV disease. The impact of these examples has been profound. First, hopes were raised by a catchy hypothesis. Then, trials were conducted, ostensibly designed to answer the question. Later, when results proved less than hoped for, innovative methods were used to suggest greater

benefit than actually occurred. In fact, the researchers, in each case, failed to completely acknowledge what their studies really showed.

There was a double fallout from this: on the one hand, many people (particularly researchers and clinicians) went about in a state of denial about the truth, by citing the "innovative" methods (e.g., responders vs. non-responders, or post hoc subgroup analysis); on the other hand, hopes once raised too high crashed, and the entire community (or most of it, particularly people living with HIV) went into a prolonged, not-yet-over state of clinical depression about clinical trials and the lack of truly effective therapies.

The Army Phase One gp160 Study. The first case study is taken from outside the ACTG, as the problems in trial design and analysis are hardly restricted to the ACTG.

If there is still someone who feels that a classical "Manhattan Project" approach might be an appropriate way to structure AIDS research, perhaps a look at the Army AIDS research program will disabuse them of that notion. In both cases, centralized, top-down research was the basic paradigm. The Manhattan Project itself had such talent on-site that a permanent kind of in vivo peer review system was in operation, but the Army AIDS program has suffered from too little outside, objective review. What was originally a small program grew to \$50 million a year (exclusive of the \$20 million FY1993 gp160 earmark; \$50M is about as large as an NIH institute with a large, but not gargantuan, AIDS program, e.g., NCRR, NHLBI or NICHD).

In line with its resistance to admitting the ubiquity of homosexuality (in and out of uniform), a 1992 Army document states that "the focus [of Army HIV studies] is on heterosexual transmission... The prevention of infection is a priority over the prevention of the progression of disease, which is a priority over the prevention of death." Vaccine research is thus a priority, and collaborative studies with the Thai army are in the works.

The Hypothesis. Therapeutic vaccination for HIV disease was first proposed by Jonas Salk in 1987. The hypothesis was that immunization with non-infectious whole virus (or, later by other sponsors, with viral proteins) would elicit new protective immunity (cell-mediated, humoral, or both) against HIV and prolong the asymptomatic state. The hypothesis had the added convenience of justifying safety and activity studies in HIV-infected patients, which seemed "safer" to many vaccine manufacturers than exposing uninfected individuals to possibly immunosuppressive antigens.

In 1987, prevailing dogma held that the viral load was very low during the asymptomatic phase of infection, that the virus was not particularly immunogenic (that is, that it was hard for the immune system to "see" HIV), and that antibodies to HIV were protective. Thus, exogenous killed-virus (or antigen fragments) might stimulate new and enhanced immunity, helping the immune system to "see" the virus, and eliciting new antibodies.

In 1993, it is clear that viral load is much higher than previously believed, even during the asymptomatic phase; that the virus is highly immunogenic and clearly recognized by both cytotoxic T lymphocytes and by antibodies, some of them neutralizing -- and it is unclear how protective antibodies are. Thus, the entire basis for the 1987 therapeutic vaccination hypothesis has been reversed by later discoveries. Yet, because drug and vaccine development remains many years behind new insights about pathogenesis, we are still locked into a 1987 hypothesis about therapeutic vaccination, assessing it using first-generation

vaccine products based on laboratory isolates of HIV which lack many features of the clinical isolates which infect people or those which later make them sick.

The Product: MicroGeneSys VaxSyn™ recombinant gp160. Even among the first-generation vaccine products, most of which are based on envelope sequences of laboratory isolates such as IIIB, the MicroGeneSys rgp160 stands out as one uniquely unlikely to elicit relevant cell-mediated or humoral responses to the virus strains that circulate in infected people.

gp160 is the protein complex, composed of gp120 and gp41, which studs the surface membrane of HIV, and which mediates binding to the CD4 receptor on target host cells. The gp160 protein itself breaks down in vivo into its two components, and is not as immunogenic together as gp120 is alone [ref.].

MGS rgp160 (like many of its sister products) is based on the canonical laboratory strain HIVHTLV-IIIB, the pesky lab strain which first invaded a French culture, then turned up in Bethesda, overtaking several other isolates along the way. IIIB is particularly well-adapted to growth in permanent T-cell lines, and differs in many important respects from HIV strains that are found in vivo.

MGS rgp160, unlike many of its sisters, is grown in an insect virus vector (baculovirus), and so it is not synthesized as are viral proteins are made in mammalian cells. Many other vaccine products are made in Chinese hamster ovary cells or other mammalian tissue.

While other vaccine products are synthesized as three-dimensional proteins in mammalian cells, with sugars attached at glycosylation sites, MGS rgp160 is synthesized as a denatured, nonglycosylated, linear peptide. Thus, the antibodies which could be stimulated by injection with MGS rgp160 are presumably antibodies to linear peptides. By contrast, the vast majority of neutralizing antibodies to HIV in vivo are to conformational (three-dimensional, non-sequential) epitopes. [While even rgp160 could elicit cytotoxic T cells (CTLs) to rgp160 peptide fragments, but the new CTLs too may fail to recognize autologous envelope sequences.

Finally, the envelope gene env itself has been programmed through millennia of escape from a vigorous host immune response to be the most variable gene of HIV (along with reverse transcriptase [RT] itself). Thus, within a single host, HIV mutates its gp160 and gp120 coat constantly to evade detection. Therefore, basing vaccines (whether prophylactic or therapeutic) on this viral element, rather than on highly conserved elements necessary for viral replication, could prove to be a disaster. Recent laboratory findings, showing that all current vaccine candidates failed to neutralize field HIV isolates, confirm this.

For all these reasons, then, one can certainly understand why John Moore, George Lewis and James Robinson opined that "In short, based on existing in vitro data, our opinion is that there could not be a worse choice from the current envelope glycoprotein vaccine candidates than MGS gp160 to stimulate at least one important arm of the immune system, the production de novo of cross-neutralizing antibodies to the V3 loop and discontinuous epitopes around the CD4-binding site." Those directing the Army AIDS research program were not deterred, however, by these considerations.

The Experiment. One disturbing but inevitable result of the urgency engendered by the AIDS crisis is that both researchers and community members tend to invest preliminary trials with more significance than they can possibly bear. This was

the case with the pilot combination therapy study ACTG 106 [see below], and it was the case with the Army's phase I gp160 study in 30 HIV-infected persons as well. In the first case, that of ACTG 106, The New England Journal of Medicine refused to publish the preliminary data; but in the second, for reasons unknown, they chose to publish results of a non-randomized, exploratory phase one study, in the American journal of record for the standard of medical care. This decision played no insignificant role in helping to fuel the subsequent hype.

In the Redfield study, 30 HIV-infected Army personnel with initial CD4 counts over 400/mm³ were given one of two injection schedules of MGS rgp160 -- on days 0, 30 and 120; or on days 0, 30, 60, 120, 150 and 180. The major endpoints of interest included new or enhanced antibody responses to gp160 sequences, as well as T cell proliferative responses to gp160.

The Analysis. The investigators divided the group into "responders" (those with new or enhanced antibody or T cell responses) and "nonresponders" (those without). Such post hoc analyses are a popular way of reading more into nonrandomized studies than is really there. In later stages of follow-up, "nonresponders" were converted to "responders" by a more frequent schedule of injections. The initial "responders" had higher CD4 cells at baseline, and allegedly sustained a slower slope of CD4 decline than the initial "nonresponders". The investigators suggested that this might be the effect of the rgp160 injections.

Several things are notable about these results, and their interpretation. First, the "new and enhanced responses" noted in the "responders were all to linear epitopes of IIIB rgp160 (in the case of the antibodies), or T cell proliferative responses to IIIB/rgp160. No results were presented demonstrating that the newly-elicited antibodies or T-cell proliferative responses applied to autologous, clinical HIV isolates from the study participants. Thus, it remains unknown whether the newly elicited immune responses did anything at all to the viruses circulating in the patients in the study. If they do not, the result is as relevant as immunizing people to classical laboratory antigens -- i.e., not relevant at all.

Secondly, the difference in the CD4 slopes among the initial "responders" and "nonresponders", if real, could have been due to differences in baseline immune competence (hence a higher baseline CD4 predicted the likelihood of being labelled "responder"), and consequent differences in the rate of decline, rather than differences due to the administration of the antigen itself. In addition, some of the study participants were taking AZT, which obviously can affect short-term CD4 levels.

Thirdly, as it emerged later, there were problems with the analysis of the data published in The New England Journal. The authors compared different "moving averages" of the CD4 slopes over time, rather than using the actual levels. This undermined the validity of Figure 4 in the New England Journal paper.

Fourthly, subsequent presentations based on longer-term follow-up from the phase one study were used to inveigle Congress into appropriating an extra \$20 million to the Army AIDS research program for a large-scale, phase three study of MGS rgp160. At the Amsterdam AIDS conference in July 1992, the Army's Dr. Redfield reported "dramatic reductions on both the levels of HIV proviral DNA... as well as full length HIV RNA when compared to natural history controls" in "the first 15 AIDS patients treated with the MicroGeneSys vaccine". Subsequently, the Army's Dr. Birx reported on viral burden reductions in a single patient at the annual vaccine conference that year.

There were several problems with this analysis. The patients were not the first fifteen patients, they were fifteen selected patients (and, as noted above, some may have been on concomitant AZT). There were no statistically significant reductions in the group of 30 patients overall. The natural history controls were not comparable to the study participants. The list goes on.

The Aftermath. The gp160 affair has been an enormous drain on everyone's energy and time. The Congress was tricked into diverting \$20 million on a useless, unnecessary trial. The AIDS community was once again subjected to the old game of bait and switch, with temporarily raised hopes doomed to be dashed once again. The NIH wasted thousands of hours of staff time convening blue ribbon panels and meetings, developing plans for a "large, simple trial" that will never be mounted. The Army has already spent \$1 million in "planning costs"! The Jackson Foundation's chief statistician, William McCarthy, resigned rather than condone the misuses to which the Army's data was being put. Army researchers were subjected to an internal investigation. Recently, Representative John Dingell suggested he might undertake a full-scale Congressional inquiry. ACT UP/New York members demonstrated at the corporate gates of MicroGeneSys in Meriden, Connecticut. In November 1993, Congress passed conference language giving the Executive Branch another six months to change its mind -- to divert the money to more scientifically justifiable AIDS research. And HIV vaccine manufacturers are arming themselves with high-powered Washington lobbyists to push their agendas in the future. A horrible precedent has been set for biomedical research in general, and for AIDS research in particular.

Often forgotten during the debate is the fact that there are almost 2,500 patients around the world already enrolled in phase II studies of rgp160 and other therapeutic vaccine candidates. If any of these trials shows anything useful, it is likely that a phase III trial will eventually be mounted. [Indeed, if such data were now available, MicroGeneSys would have no trouble raising capital for such a study from its partners, Wyeth-Ayerst and American Home Products.] But it would simply be a waste of money to conduct a phase III trial now -- it would amount to throwing up our hands in the air and saying who needs markers, who needs preliminary signs of activity, we might as well throw everything into large simple trials.

Data are wilfully misinterpreted. Time is wasted. Money is wasted. Lobbyists and legislators set research priorities. Hopes are dashed. We are trapped in the outdated clinical research paradigms of the mid-1980s -- not only in vaccine research, but also in antiretroviral research, with its plethora of me-too nucleoside analogues and their non-nucleoside reverse-transcriptase inhibiting cousins -- while pathogenesis research moves on into uncharted territory, its promised clinical payoff cruelly far off to anyone now living with HIV.

A Second Case Study: Combination Nucleoside Therapy

The Hypothesis. In general terms, as we've heard ad nauseam for years, certain therapeutic combinations are more effective than single agents for chronic diseases such as tuberculosis or cancer. In addition, certain combination antibiotics [with different molecular targets] are useful clinically, e.g., trimethoprim-sulfamethoxazole. Two nucleoside analogues can be more powerful than one in vitro; two antiretrovirals generally (even clinically ineffective ones, such as dextran sulfate or CD4) can be "synergistic" in vitro. Finally, preliminary data from pilot combination studies (e.g., ACTG 047, 050, 106, 143; as well as Yarchoan's NCI study, Collier's AZT/ddI study, and Schooley's viral resistance study for Burroughs-Wellcome) suggested (but did not prove) that

combination nucleoside therapy might be clinically superior to monotherapy, at least temporarily.

The Agents. "Until there's a cure, there's Retrovir". When ACTG 155 was designed, everyone with $CD4 < 500/mm^3$ was recommended to go on AZT for an indefinite period. Therefore, a clinical study of combination therapy could have addressed two populations: 1) a first-line population, randomized to AZT alone vs. AZT and another nucleoside; 2) a second-line population, randomized to continue AZT, add another ddN, or switch to the new ddN. The first-line population had a greater chance of demonstrating efficacy (because they would not yet be resistant to AZT), while the second-line population was more relevant clinically (because everyone was supposed to be on AZT already). The designers of ACTG 155 chose the second population, and for some reasons unknown chose HIVIDTM brand ddC (zalcitabine) as the relevant second nucleoside. One wonders why they didn't use both available nucleosides, ddC and ddI, as the British Medical Research Council (MRC) did in designing its Delta trial (AZT vs. AZT/ddI vs. AZT/ddC as first-line therapy, N=1717), since it was already evident in 1990 that ddC as a monotherapy was definitely no better, and quite possibly worse, than ddI alone. In this case, it appears that the ACTG leadership succumbed to the resistance of the corporate sponsors to a head-to-head comparison of the "new" nucleosides.

ACTG 155 was riddled with investigator bias and sponsor-disseminated hype from the very beginning. The investigators were "sure" combination therapy was better. It was used as a recruiting tool and then, when the guaranteed crossover to combo was proffered, as a bribe to patients for staying in the trial on blinded therapy until they reached a primary endpoint.

Indeed, as early as August 1990, at a meeting with ACT UP, Roche's Dr. Whaijen Soo told us: "I'd like to confide a remark of Margaret's... The difference she sees among people on the ddC/AZT combination [in ACTG 106] is a big difference - - like the difference between people on AZT versus placebo in the original trial. She can almost tell them apart by looking at their behavior." Say that around the country for three years and you can create a huge demand for an unproved drug.

Thus it is no wonder that ACTG 155 was one of the fastest-enrolling ACTG trials ever, and that demand for the regimen far outstripped the capacity of the trial. Roche continued to delay implementing a true expanded access program, and underground use of ddC, procured from chemical supply houses in the USA and abroad, burgeoned. In early 1992, FDA analyses showed that some batches of this underground ddC contained unacceptable dose variations, and the supply was cut off. Only weeks before the FDA hearing on the ddC NDA, Roche finally opened a true expanded access program for ddC used in combination with AZT. Nonetheless, this use obtained only the slenderest of approvals at the FDA, being licensed under the new conditional "accelerated approval" regulations based on a 150-patient resistance study sponsored by Burroughs-Wellcome (not by either the ACTG or by Roche!). Understandably, then, all concerned were eager for the hoped-for clinical confirmation from ACTG 155 of combination therapy's long-touted benefit.

The Experiment. Principal investigator Margaret Fischl, of the University of Miami School of Medicine, who will become chair of the ACTG Primary Infection Committee next year, presented the ACTG 155, the first study of combination therapy to use clinical endpoints, to the FDA advisory committee on September 20, 1993:

We did a study to ask whether [1,001] patients on long term (over six months) of AZT therapy, and symptomatic with $CD4 < 300$ or asymptomatic with $CD4 < 200$, would be

better off adding ddC, switching to ddC monotherapy, or staying on AZT. We stratified at baseline for duration of AZT therapy, PCP prophylaxis, etc. Partway through the study we added a crossover option to open-label combination therapy for patients who reached a primary endpoint [new AIDS-defining condition; presumably those who died were not offered combination therapy]... At the time we designed the study we still thought we might see a difference between combination therapy and monotherapy... In retrospect this was probably the worst population to study... We're presenting a non-traditional analysis for your information.

The Analysis. In Berlin, at the first public presentation of the data from ACTG 155, Dr. Fischl hailed it as "the first study to demonstrate the clinical benefit of combination therapy." While NIAID hurriedly redrafted a press release to eliminate such remarks, she continued to maintain this line throughout the Berlin conference, and at the ACTG Retreat during July. By September, the 155 team had gotten its act together for the FDA hearing, and solemnly admitted the null overall results before presenting their now famous "post hoc subgroup trend analysis" based on three baseline CD4 strata (>150, 50-150, <50). Later in the day, Roche scientists produced their preferred two subgroup analysis (>/<100), claiming that the benefit "appeared to change direction" around a baseline CD4 of 100/mm³. Either way, the results suggested a benefit to combination therapy at higher CD4 levels, and one to monotherapy (especially AZT monotherapy) at lower ones. Overall, combination therapy was 50% more toxic than either monotherapy, but no more effective.

The development of ddC was accompanied from the start by inappropriate exaggerations of its benefit. Last year, Roche decided to ignore the bad news of ACTG 114 and the inconclusive results of ACTG 119, to tout the improbable ACTG 106 results as showing more than they did, and to apply for approval in the absence of any evidence of benefit. This year, they continued that pattern, distorting the disappointing results of CPCRA 002 and ACTG 155 into a shaky package, apparently fully confident that the FDA, its advisory committee, and the AIDS community will continue to embrace ddC in spite of the failure of all the previous hopes.

In ACTG 155, ddC monotherapy patients had the highest proportion of clinical events, and ddC combination patients had the highest proportion of deaths. But overall the study failed to show statistically significant differences among the three arms.

The much-touted "subgroup trend analysis" shows very little. The trial wasn't stratified by CD4 counts at the outset. The high baseline CD4 group (150-300) -- the one with the claimed "benefit" of combination therapy -- was exactly the arm with the fewest clinical events, and thus the least power. The high CD4 subgroup had just 15.5% of the clinical endpoints, and only 8% of the deaths. Thus, this subgroup lacked the number of events to provide enough power to clearly tell us anything -- especially when those groups with the most endpoints tell us quite a different story.

The people who entered in the lower two CD4 subgroups (<150) had 85% of the progressions and 92% of the deaths. This is where the study's power is. Moreover, in the lowest CD4 subgroup, ddC alone was the worst for progression, and combination therapy was worst for death. The sponsor handles this apparent contradiction with the statement that "the treatment differences reversed directions at a baseline CD4 count of approximately 100 cells/mm³".

How can one accept the putative "benefit" of combination therapy in people with CD4>150 if one ignores the potential "harm" of ddC monotherapy or combination therapy in people with CD4<50?

ACTG 155 is a wash. It's time to rethink how NIH and industry design and analyze efficacy studies, and how FDA regulates them. An NIH statistician recently told me that all the ACTG phase II/III studies are underpowered. But NIAID has too few statisticians to review the protocols before they start. It's as if NASA didn't have enough engineers to make sure their rockets could fly. [Indeed, lately it looks like they don't.] But when a rocket blows up, at least everyone knows about it. When a drug trial crashes, the sponsor goes ahead and demands an NDA. In any case, the benefits of these drugs are so marginal that they may only eventually be revealed by much larger studies.

In the end, ACTG 106 showed very little, and ACTG 155 showed us no difference at a much higher resolution. Yet the team wrote to the patients, "In summary, the study showed that patients with advanced HIV disease who have received long-term ZDV therapy and who have CD4 cell counts > 150 cells/mm³ have a better outcome if they receive combination therapy with ZDV and ddC than if they continue on ZDV therapy..." The participants who survived the trial deserved to hear the full truth, but the investigators denied them this courtesy. But then again, if researchers can tell who's on what regimen, and which regimen is a winner, by obvious behavioral clues, why do trials at all?

The Implications. Once again, as with the Army's phase one gp160 study, data are wilfully misinterpreted. Time is wasted. Money is wasted. Hopes are dashed. We are trapped by assumptions which have been refuted by the Concorde study and by ACTG 155.

And all our follow-up trials -- ACTG 175, 193, and the phase two trials of the NNRTIs -- are based on the outdated, or at least far from secure, notion that preliminary CD4 changes predict clinical benefit. None of these studies (with the exception of 193, which is a useless travesty) is powered to show which regimen is more beneficial clinically.

Increasingly -- in ACTG 229, 241, 244 and 261 -- control arms in antiretroviral studies use combination AZT/ddC, or AZT/ddI, as though these had been clinically validated as the standard of care, when, after ACTG 155 and until something better comes in, THE OPPOSITE IS THE CASE: Combination therapy with AZT/ddC in the 155 population is 50% more toxic and no more effective than monotherapy with AZT alone. Using combination regimens as pseudo-control arms in current studies will only further confound their results, leaving us farther than ever from providing people with HIV and their providers with clear, relevant, useful treatment information.

A Third Case Study: Convergent Combination Chemotherapy

The Hypothesis. See notes on the generic rationale for combination antimicrobial therapy in the previous section. In this case, the usual, and rather pedestrian insight of combining three antiretroviral agents in vitro to produce "synergy" was originally presented, without undue attention, at the Amsterdam AIDS meeting in 1992. Later these results were dressed up with a catchy theory about the "use of evolutionary limitations of HIV-1 multidrug resistance to optimize therapy" and published this February in Nature. In the meantime, the ACTG Executive Committee approved ACTG 241, a phase II study of the "convergent" regimen of AZT/ddI/Nevirapine compared to AZT/ddI alone, using the standard Combination

Therapies Working Group master protocol format of 100 patients per arm with CD4 counts as the primary outcome measure. So far, all seemed ordinary.

But February 1993 was not an ordinary time. A new administration was in power, one that seemed more interested than the previous ones in AIDS research. A bill was being debated on the Senate floor which would radically reform the structure of the NIH AIDS research program. Some NIH officials were waging a furious campaign to defeat this initiative, using professional societies (such as ASM and AAMC) and a minority of academic researchers as proxies. Meanwhile, Harvard was circulating a plan to take over AIDS research.

So the Nature paper, for a number of unrelated reasons, was subjected to an unusually bright glare of publicity. Massachusetts General Hospital issued a press release describing the discovery as a potential "Achilles heel" of HIV [ref.]. The mass media went wild. The story was front-page news across the country. Yung-Kang Chow, the medical student who was the primary author, was showcased as ABC News' "Man of the Week"; his concept of drug-inducible evolutionary limits to HIV's ability to mutate was presented as a scientific breakthrough on the lines of Archimedes' "Eureka!" Beleaguered OAR and NIAID Director Anthony S. Fauci appeared on television news playing the drug (presumably AZT/ddI/Nevirapine, the new trinity), saying, "You're the virus. I'm the drug. I'm giving you two choices: either I kill you, or I make you mutate yourself out of existence." NIAID put out a press advisory listing the ten sites where ACTG 241 would take place. One week later, NIAID issued a second advisory noting that six sites were being added. Quickly the Primary Infection Core Committee recommended that the study size be doubled, to 400 patients. No scientific rationale for expanding the study was presented, and no one from the protocol team objected. An AZT/ddI/Nevirapine arm was added to the ACTG 193 study, in an effort to make this pointless and unnecessarily complicated study more "attractive" (if yet more complicated) to potential participants, who must have CD4<50/mm³. The research team began to claim it had begun to discourage media interest, but not until several mass-market magazines had already published photos of Dr. Chow holding up vials labelled "AZT", "ddI", "Nevirapine", like some new magic elixir. After stoking the flames with its press release, Mass General spokesman Martin Bander charged that the media "revisited the subject too many times and gave it too much prominence." What did he expect? Within weeks, Harvard Medical School received a multi-million dollar gift [from whom? -- check].

Privately, many were skeptical. One researcher involved with the work privately told an ACTG colleague the night before the story broke not to get too excited. A DAIDS executive quietly asked a reporter from a major scientific journal not to hype the story. But it was too late.

Phyllis Sharpe, a 41-year-old woman from the Bronx living with HIV with an infected daughter, told The New York Times that "To me, this is really the first big hope. I'm scared to even talk about it. I'm just praying within that this is really true." Sites enrolling patients into ACTG 241 were besieged with people clamoring to get onto the waiting list. One volunteer later said, "I was surprised at how jealous many of my fellow members of the support group were when I got into the trial [ACTG 241] - 1 of only 25 people accepted out of more than 400 applicants". Boehringer Ingelheim, manufacturer of Nevirapine, was besieged with inquiries from the community, and a team toured the country meeting with treatment activists during March and April.

The Agent. What was all the fuss about? Nevirapine was just one of a number of second-generation reverse-transcriptase inhibitors (the so-called non-nucleoside

RTIs, or NNRTIs), along with Merck's L-697,661 ('Pyridinone'), Upjohn's U-87 ('Ateviridine') and U-90 ('Delavirdine'), which several drug companies discovered in the late 1980s during expansion of routine HIV RT inhibition assays to non-nucleoside compounds. These compounds had decent in vitro antiretroviral activity and entered phase I studies within and without the ACTG, alone and in combination with AZT.

The usual roller-coaster ride of expectations within the community started to take off. First euphoria (as though another generation of RT inhibitors were just what everyone needed -- schematic pictures of Nevirapine in action were even published in Artforum!), then despair (when the first reports of rapid viral resistance to Nevirapine and L-697,661 were made in December 1991).

The ACTG kept plugging away, slowly completing the 168 trial, which suggested that higher doses of Nevirapine might delay the onset of resistance, reigniting interest in the 164 monotherapy study, for which enthusiasm was now understandably somewhat diminished in the community. In addition, ACTG 208 was deployed to address whether giving the drug to people with CD4>500/mm3 might delay the onset of resistance, and 20 patients enrolled. Nothing truly untoward occurred before February 1993.

The Experiments. The Nature paper discussed a number of in vitro experiments combining various nucleosides and non-nucleosides, including AZT with ddI and Nevirapine, Pyridinone (Merck's L-697,661), or Foscarnet [thankfully, most coverage did not focus on the third agent, which is licensed, highly toxic, and outrageously expensive]. The work was carried out by Harvard medical student Yung-Kang Chow and colleagues under the supervision of virologist Richard D'Aquila and Martin Hirsch, current chair of the ACTG Primary Infection Committee.

1. Proviral DNA Mutants. The team tried to construct "mutant provirus DNAs" by introducing known resistance-associated point mutations into DNA plasmids and then transfecting these "into COS-7 cells...". Then they assessed whether the transfected mutant proviral DNA could infect T cells -- which it couldn't. They took this as proof that the drug-associated mutations they'd introduced were responsible for the failure to infect, assuming that therefore they had constructed lethal mutations in reverse transcriptase.

2. Convergent Therapy in A Laboratory T-Cell Line. Secondly, the administration of the AZT/ddI/Pyridinone combination [Nevirapine wasn't mentioned here] to established HIV-infected T cell cultures was said to eliminate viral DNA from culture at 21 days, to eliminate p24 antigen by 35 days, and to continually suppress virus for 101 days, including 45 days after removal of the drugs. The HIV strain used was not stated.

3. Convergent Therapy in A Clinical HIV Isolate. Thirdly, "after ten PBMC passages of a clinical HIV-1 isolate with AZT- and ddI-selected mutations in AZT+ddI+Nevirapine we were unable to select for viruses triply resistant to AZT + ddI + Nevirapine...". On reading this passage one immediately wondered what would happen after 20 or 30 passages -- which we were to learn by June.

The Aftermath. While the mass media exulted, and people with HIV clamored to enter the trials, many virologists were concerned. Even the "News and Views" piece, written by virologist Douglas Richman, which accompanied the February Nature article, commented that "HIV is a highly mutable and wily opponent, however, and that it may be able to avoid checkmate with a large array of alternative moves (mutations) is a real possibility. Both in vitro and clinical

data (unpublished observations) suggest that this may be the case." Others were even more critical the April issue of The Journal of NIH Research: "The experiments are technically flawed... The combination of mutations is not representative of the drug-resistant genotypes we see in the clinic... They did not prove in any rigorous way that the mutations they built into the reverse transcriptase lead to a defective enzyme," said Wellcome's Brendan Larder, co-discoverer of HIV's ability to mutate into an AZT-resistant phenotype. "The data are shaky. It is surprising Nature published the paper... [the Harvard team] are dealing with a negative result, and negative results are notoriously difficult to interpret," said the Pasteur Institute's Simon Wain-Hobson. Larder went on to inform the reporter from The Journal of NIH Research that "We have made a virus that contains mutations [associated with] resistance to all three drugs. These viruses are viable. They grow, replicate, and infect cells."

The team reacted defensively to the skepticism of its peers, retorting that "The finding of such [a triply-resistant] virus in vitro does not diminish our interest in evaluating the combination of AZT, ddI, and an NNRTI in clinical trials." Fine. But do then do not commence the trials in a blaze of publicity based on a hypothesis which you are already moving away from. The team went on to address "contentions that technical artifacts might have occurred in constructing or transfecting our mutants and that the mutations we studied are not 'representative'", stating that "We did not use these experiments to support the hypothesis that combinations of mutations that are incompatible with virus replication can limit multiple drug resistance". What were they suggesting, then? That was exactly the point of Chow's "Eureka!" experience so widely touted in the press. They then defended the rigor of their proof that their induced RT mutations "led to a defective enzyme," noting that another manuscript was in preparation documenting this, and said that "undetected cloning anomalies are a far-fetched explanation for our results..."

The team may later have come to regret its tone of infallibility. As early as the International AIDS conference in Berlin during June -- just as its letter was appearing in The Journal of NIH Research, Yung-Kang Chow announced at a plenary session that triply-resistant viruses did, in fact, emerge after between 20-30 in vitro passages. Larder's team from Wellcome and Emini's from Merck also reported their success in culturing multiply-resistant HIV that month at the Second International HIV Drug-Resistance Workshop in Noordwijk, the Netherlands.

Apparently the Wellcome and Merck team's results motivated the Harvard team to offer to exchange mutants and see what the discrepancy was. Before sending off the Harvard mutants, Hirsch thought it would be prudent to check their sequences, and in doing so found that the proviral sequences had extra, previously unnoticed nucleotides. The drug-resistance point mutations were not the ones which made the mutants nonviable. The unnoted nucleotide substitutions coded for nonsense protein, and the provirus could not replicate at all. A retraction letter was drafted and sent to Nature.

Shortly thereafter, The New York Times broke the story in public. In this case, the newspaper addressed the troubling underlying issues with more frankness than would the scientific press. Chow was forbidden to speak to reporters. The Times reporter noted the usual rationale for combination therapy, but said "scientists have not yet come up with documented evidence that any combination of drugs is beneficial for HIV or any other viral infection." Despite the retraction, "Dr. Hirsch and Dr. Anthony S. Fauci... said in separate interviews yesterday that the flawed laboratory study was not a reason to stop the combination drug trials. The concept of combination therapy remains valid, they said, based on a widespread belief [emphasis added] among experts...". Sometimes the belief in

combination therapy seems to overtake the need for a careful exploration of how best to study it. Having failed to show that two drugs are better than one, we are now comparing three drugs to two, as though the issue were settled. Moreover, the choice of which agent(s) to add seems arbitrary. Why not establish optimal two-drug regimens (if there are any), then base three-drug regimens on the outcome of careful, replicated laboratory and pilot clinical work? To his credit, Dr. Hirsch made a rapid public retraction. But it was full steam ahead on ACTG 241; "the confused volunteers are being asked to plunge ahead and subscribe to the belief that three drugs may be better than one or two". Many will surely do so, since they have been so effectively programmed by years of pro-combination propaganda from the leading researchers in the field. "The error also revealed a surprising degree of sloppiness in the quality of bench research conducted by researchers at one of the most prestigious hospitals and medical schools in the world... Mr. Chow published without repeating crucial steps to verify ambiguous data... The team misread laboratory data and failed to detect a fifth mutation in the virus... This incident adds to a list of the times when the much-heralded peer review system failed to detect the flawed research it was designed to weed out... Reflecting its initial enthusiasm, the Harvard team generated publicity by issuing a news release saying its research might have found the 'Achilles heel' of HIV. But a few days later, the Harvard team stopped talking to reporters because the publicity got out of hand, a reaction that the team naively says it never anticipated."

On August 5, the editors of Nature announced that "Chow et al. have sent a retraction of part of their work... The authors admit that the discrepancies between their data and those of their critics (some of whom had earlier submitted them to Nature) can be explained by four [emphasis added] previously unnoticed mutations in the reverse transcriptase of one of their HIV-1 clones. They go on to say that this development does not affect the accuracy of their other data, but it does invalidate their claim to have proved that multidrug therapy will be effective by avoiding drug resistance." The editors go on to congratulate themselves for publishing the correction quickly, sadly noting that "the publicity does not help". They did not address whether their own decision to publish had helped to fuel the publicity (Nature rarely publishes drug development papers, just as The New England Journal rarely publishes phase I studies like the Army gp160 trial).

In its retraction letter to Nature, the study team confessed that the explanation which it had previously found "far-fetched," that is, "undetected cloning anomalies," is actually correct: "The clones... of mutant 4 had unintended mutations in addition to the intended mutations in reverse transcriptase codons 74, 103, 215 and 219... We believe that all the other data in our report are correct." They go on to note that they, too, "have also been able to select for triply resistant viruses similar to those described by Larder et al." Accompanying their letter is one from scientists at Merck Research Laboratories describing their successful efforts to culture HIV-1 variants resistant to AZT, ddI, L-697,661 and/or BI-RG-587 [Nevirapine], thus suggesting that quadruply resistant mutants are feasible, and suggesting that "such variants are likely to be selected during multiple therapy" By September 30, Larder and colleagues had reported that "HIV-1 co-resistant to AZT, ddI and the NNRTI Nevirapine can be readily selected in cell culture starting with dual AZT- and ddI-resistant virus. We found no evidence for 'replication incompatible' combinations of resistance mutations." Passage of one construct "concurrently in AZT, ddI and Nevirapine led to the rapid [emphasis added] development of Nevirapine resistance, and, curiously, an actual increase [emphasis added] in AZT resistance." They also constructed the mutant which the Harvard team stated it had made in the February paper, and found that "recombinant reverse

transcriptase containing these mutations... was virtually as active as the wild-type enzyme." Having demolished virtually the entire ensemble of claims made in February, the Wellcome team went on to determine that an HIV-1 variant triply resistant to ddI, Nevirapine and the nucleoside analogue FTC had actually regained susceptibility to AZT! This, if it could be replicated in the clinic, would be a perfect strategy for returning people, several years, drugs and thousands of dollars later, to the AZT-susceptible state at which they began therapy -- and a boon for Burroughs-Wellcome. But the deeper lesson is that HIV quite rapidly can evade any regimen composed exclusively of reverse transcriptase inhibitors. Perhaps RT alone is the wrong target.

The Implications. Larry Altman of The New York Times made a key point in his wrap-up on the whole affair: "Anyone can come up with a theory about the treatment of a disease. But there is not enough time, money and expertise, let alone willing volunteers, to test every new theory." Not all theories are equally meritorious. Nor are all investigators equally good, or all journals equally reputable. But when a highly-regarded, powerful investigator publishes something catchy in a top-ranked journal, it may not be subjected to the same skepticism with which most new and unexpected results are received. Within the ACTG, there is no true peer review in the Primary Infection Committee, and too little oversight by the Executive Committee. Therefore, it was smooth sailing from the first for a catchy hypothesis that more careful work would have doomed from the outset.

But the impetus for studying combination therapies in the clinic is not deterred even for a moment by such retrospective setbacks. The trial is underway, is it not? -- and it's twice as large as it needs to be -- who needs a hypothesis now? The machine is working! Who cares if its output is unreadable, or that people with HIV have poor information on which to base their treatment decisions, or that other approaches will not be studied because the machine is clogged with nucleosides and NNRTIs? The ACTG has already programmed all its trials for 1994. New approaches cannot be studied unless old ones are scaled back. Each new antiretroviral trial seems to show less and less obvious clinical benefit. Rather than assessing why we can't even answer clinically relevant questions for antiretroviral therapy in the ACTG, those running it simply pile on more reverse transcriptase inhibitors (RTIs), cross their fingers and hope for the best. Statistically, if they run twenty combination therapy trials, at least one will break "positive" simply by chance. And that one statistical fluke will then be taken and broadcast around the world as proof of the "efficacy" of combination therapy.

*

Follies of 1993, or How Did We Get Into This Mess

(And How Do We Get Out?)

The Omni-Drug Revolution. In the campaigns against disease, researchers have attained some success by utilizing two or more drugs together. Recently, this approach has led to the development of omni-drug therapy, in which hundreds of different drugs are administered together. Proponents argue that the body's immune system selects only those drugs that are beneficial, and renders the other drugs inert. Opponents argue that the claim is farfetched and extremely dangerous. Omni-drug therapy has received much favorable attention in the general press. Patient support groups for several diseases are applying political pressure to allow experimenters to skip preliminary laboratory trials with animals and proceed directly to human clinical trials.

-- The Journal of Irreproducible Results September/October 1993

The report of the recent Ad Hoc Scientific Review of the ACTG questioned "whether many of the current large efforts (such as the nevirapine trials) are scientifically sound or whether the trial design will answer the study's research question."

Having failed to prove that two nucleosides are better than one, the potentates of the ACTG primary infection effort are busy designing multiple phase 2 studies, incompletely controlled with ill-defined "control" arms like AZT/ddI or AZT/ddC, comparing three drugs (2 nucleosides plus an NNRTI) with two. It seems that the "Omni-Drug Revolution" is well underway in the ACTG. This would be a windfall for industry.

Samuel Broder once commented that with AIDS, as has already occurred in cancer, we are in danger of being locked into mediocre therapeutic regimens. So far, the clinical data on combination therapy with AZT/ddC indicates that, overall, it is 50% more toxic than monotherapy, and no more effective. So participants on the control arms of these new studies (ACTG 229, 241, 244, 261) may be being exposed to useless toxicity and inconvenience without corresponding benefit, and those "lucky" enough to receive triple-drug therapy are exposed to the toxicities (rash, etc.) and the ill-understood, transient anti-HIV activities of the NNRTIs.

Meanwhile, those evaluating the current "state of the art" in terms of antiretroviral standard of care, have apparently moved away from reliance on ACTG studies, and are framing the question of how best to use current anti-HIV drugs in terms of "patient choice" and "the art, not just the science, of medicine." The new NIAID-sanctioned guidelines contain a total of twelve optional regimens based on stage of disease, and can be summed up basically in two words: "Anything Goes." It is as though the research establishment is simply giving up on the need to discover how best to use current therapies. This implies equal pessimism about how to define optimal regimens with new therapies. And the hypotheses now being studied in the ACTG are unlikely to provide clear, rapid answers to these questions.

The [Primary Infection] Committee appears to lack specific guiding principles by which to set its priorities and guide its future work. For example, should they test the best available drugs, drugs that are hard for industry to test, or orphan drugs; concentrate on trials that set standards of care; be involved in the early evaluation of every drug, or in fine tuning earlier established data? Once the overriding principles are clear, the committee could then define and follow its priorities. In addition, the committee appears to have few criteria for evaluating what many on the review panel view as junk drugs (dextran sulfates, nevirapines, or therapeutic vaccines).

ACTG 175 was designed in the heyday of optimism about surrogate markers. Primary Infection chair Douglas Richman once commented that "ACTG 155 will be the last PI study to use clinical markers." 175's designers supposed that they were extending the results of ACTG 019 vis-a-vis early intervention, by comparing AZT to ddI monotherapy, and to AZT/ddI and AZT/ddC in combination, in a similar population. Unlike 019, however, the endpoint of interest was to be CD4 cell changes. The results from 019 had never been analyzed to show whether AZT-induced CD4 changes were in fact predictive of benefit. And most of the endpoints in 019 were PCP ones, because PCP prophylaxis was strongly discouraged during that study. Now, with virtually everyone in ACTG studies on prophylaxis when their CD4 drops below 200, there would be very few PCP endpoints. Thus, CD4

changes were regarded as a more practical and "affordable" option than waiting several years longer for clinically significant differences to show up. After Concorde's results were released, the ACTG 175 team added clinical endpoints to the primary objectives, but there is only enough power to compare "immediate vs. deferred combination therapy" rather than to compare the 4 arms individually to each other.

The controversy during the planning of ACTG 175 divided activists and scientists alike. Some thought the trial a wholesale waste of time ("You don't combine 2 drugs with the same mode of activity -- like ampicillin and penicillin," one researcher told me), while others thought it was a prudent and sensible step forward ("A Spanking for T+D," Bob Huff, September 1991). The original protocol chair, Deborah Cotton, stepped down rather than lead a trial which she thought ill-designed. There was no forum within the ACTG for the essential scientific differences to be openly debated. The protocol team slimmed down the study to get it through the Executive Committee at an N of 2,100 and an estimated cost of \$16 million. Later the trial ballooned in size to over 3,000 and was extended in duration for another year; thus the ultimate cost will approach \$25 million or more, including the nested virology substudies. The trial appears to be too large for the ACTG, and too small to answer its question definitively. First of all, the rate of clinical progression in the original 019 population was tiny -- about 3% overall, 4% on placebo and 2% on AZT. Nonetheless the study was used as a basis for recommending that the whole 019 population <500 be placed on AZT, though 97% of them wouldn't even progress within the 19 months median follow-up from 019. Now three other regimens were being compared with AZT (2% progression in 19 months) with concomitant use of PCP prophylaxis. The rate of clinical progression will be minuscule. The other regimens would have to be dramatically better or worse than AZT if 175 was to show a difference, and given what we know of these drugs, that's highly unlikely. In commenting on the amount of time that the gp160 controversy took in 1992-93, an NIH employee said "\$20 million? That's nothing! We throw that kind of money away around here!" The ACTG 175 story makes that all too abundantly clear.

ACTG 193 is a mess of a study. Rather than allowing people with late-stage disease in to the studies of the newer antiretrovirals, the ACTG designed a "salvage study" using agents that had already failed in the study population. People with <50 CD4 cells, most of whom have already been on the three available nucleosides, are randomized to AZT with ddI or ddC, to an alternating regimen, or to triple drug therapy, and followed until death. In the subgroup analysis from ACTG 155, combination therapy was more toxic and less effective than AZT monotherapy in the subgroup with CD4<50/mm³. More recently, people with low CD4 levels have been excluded from the tat inhibitor trial (ACTG 213) and the new Upjohn U-90 studies (NIH intramural study and ACTG 261). After excluding the most endangered population from its studies, the investigators go on to lament the lack of clinical endpoints in their "pivotal" trials using healthier populations. The solution is staring them straight in the face -- stop excluding people with low CD4 counts!

The [Primary Infection] committee's reliance on a single investigator's consulting role with industry to bring in new drugs is a potential problem source. Lack of full disclosure of relationships to industry erodes credibility.

ACTG 229, the Roche protease inhibitor study, was designed in an egregious and secretive manner. The company refused to provide preliminary anti-HIV activity data from its three European trials to the Primary Infection committee or the Combination Therapies Working Group. We were asked to vote blind on an expensive combination study. Roche also insisted that AZT/ddC be the "control" arm, though

this lacked validation either vis-a-vis virology, CD4 changes, or clinical outcomes. If the protease/AZT arm appears "equivalent" to the ddC/AZT arm, does that mean the protease inhibitor is "equivalent" to ddC? Hardly. Roche later showed the investigators a few shreds of activity data but again refused to allow the working group or the full committee to evaluate the results. Here, ACTG resources were being manipulated by a pharmaceutical company for its own purposes, and ACTG investigators were so eager to get their hands on a new class of compounds that they waived their duty to evaluate the concept sheet critically. Still later the trial was extended in duration, again piling on costs, thus limiting the number of new studies for 1994. Apparently Roche, like many pharmaceutical companies, still believes that CD4 changes in small phase 2 studies are adequate for approval of new drugs, even new classes of compounds.

Rather than comparing critically the various non-nucleoside reverse transcriptase inhibitors, and choosing the one(s) which had the best activity, safety, immunological response, and resistance profile before going into expensive phase 2 studies, the primary infection committee simply rushed with a combination based on a laboratory study later proved to be inaccurate, and mounted not one but two expensive, high-tech trials (ACTG 241 and 244) using just one of the available possible combinations -- AZT/ddI/Nevirapine. It would have made more sense to use a comparative format in which other NNRTIs were compared with Nevirapine in vitro and in small pilot clinical studies before rushing forward with unsubstantiated hypotheses about a single regimen. ACTG 244 adds another twist, which is to add a therapy (ddI to the AZT monotherapy arm, Nevirapine to the AZT/ddI arm) when patients develop an AZT-resistance mutation at codon 215 in their HIV RT RNA. Again, recent developments in the field were not incorporated into this design. In Berlin, virologist Victoria Johnson of Birmingham, Alabama, presented analyses of the development of AZT resistance from participants in ACTG 116B/117. Development of AZT resistance was an independent predictor of progression, and those who developed resistance to AZT did not appear to benefit from switching to ddI. These results might undermine the approach being taken in ACTG 244, but the Primary Infection Committee was not permitted to evaluate the hypothesis, but simply the design of the study, as though those two things were so easily separable.

Finally, the ACTG 261 study design replicates the unreflective clutter and lack of coherence found in ACTG 229. Upjohn's U-90 (Delavirdine) will be combined with AZT, with ddI, with both, and compared with AZT/ddI together, in a 456-patient study using CD4 changes as the primary endpoint. Again, AZT/ddI remains an incompletely validated "control" arm. Again, surrogate markers are privileged, while clinical endpoints are sidelined, along with people with <100 CD4 cells, who are excluded.

So far, then, it appears that the leading researchers in the ACTG Primary Infection Committee are incapable of framing a hypothesis in a clear and simple manner. They sometimes base large and expensive studies on anticipated confirmation of methods (e.g., CD4 changes predicting clinical benefit) which are not ultimately confirmed. Thus, every trial builds upon the unproved assumptions which helped to shape previous trials. When previous assumptions are disproved, the subsequent trials are not modified or redesigned in light of the new knowledge. The entire edifice appears to be a house of cards built upon a foundation of quicksand. The cost of the trials, in aggregate, will total hundreds of millions of dollars, and yet we are not much further with the standard of care than the helpful motto "Anything Goes".

We have not devoted much attention to how we actually do these trials...

-- Daniel F. Hoth, 11.20.89

Although it is difficult to prove, treatment of HIV has likely resulted in prolonging the life of HIV-infected individuals and improving their quality of life.

-- Margaret I. Johnson + Daniel F. Hoth 5.28.93

As I said in the introduction, the results of recent clinical trials have been disappointing. They have been so in two quite distinct ways.

First, the trial results are disappointing because there is not a large benefit associated with currently available antiretroviral drugs. That is, the drugs are disappointing and inadequate.

Secondly, they are disappointing because they do not reliably answer the question of what magnitude of benefit might exist with these drugs. That is, the trials are disappointing and inadequate.

Many people still do not understand that, while Concorde and ACTG 155 failed to demonstrate clear clinical benefits, this does not preclude clinically meaningful, but less dramatic, differences from existing; the trials were simply unable to provide a high enough degree of resolution.

If we choose, as ACTG investigators and SDAC statisticians would prefer, to emphasize the first sense in which the trials have been disappointing, e.g., that the drugs don't do very much, then presumably we may correct this by finding fabulous new agents -- perhaps the omni-drug revolution beckons here. The same-sized trials we have now, with similar designs, would be adequate for finding drugs with dramatically superior activity to anything we now have. The problem is that such breakthrough discoveries are infrequent and impossible to predict. The normal work of clinical research will be in testing agents with modest efficacy at best, refining what's already available, and improving the standard of care. The current SDAC/ACTG paradigm is inadequate to this task. In ACTG terms, a "big" trial is one with over 1,000 patients. Yet to detect differences of say 15-30% between treatment arms, an early intervention study might have to have several thousand patients.

The Primary Infection Committee should assess whether it is attempting to field too many protocols... Many protocols fail; mid-size trials often do not have sufficient power to answer questions; and studies do not seem to build well on one another. Although the committee has a capacity of 3,000 patients per year, its concepts require 9,000 patients.

Perhaps the ACTG is simply the wrong place to do truly definitive phase 3/4 studies, at least for antiretroviral drugs. As long ago as May 1990, while ACT UP was "storming the NIH," a coalition of AIDS advocates stated that:

We urge that the ACTG concentrate its efforts on conducting technology-intensive Phase I/II studies with a greater variety of new compounds than in the past. Larger Phase II, III and post-marketing studies should generally be conducted by community-based groups in conjunction with pharmaceutical company sponsors...

While the new NIAID HIV/AIDS Research Agenda notes that "treatment research is supported through four complementary clinical trials mechanisms," the institute has never made an effort to create a truly coordinated treatment research infrastructure, strengthening the unique advantages of each mechanism (ACTG,

CPCRA, DATRI and NIH Clinical Center) and fostering collaboration. The ACTG Opportunistic Infections Committee has undertaken difficult, but encouraging collaborative studies (TB prophylaxis, ACTG 177; and MAI prophylaxis, ACTG 196) with the CPCRA, but the Primary Infection Committee has done nothing of the sort.

NIAID's Research Agenda also fails to address the unresolved issues about the standard of care for HIV disease, and hence it is no surprise that the "Primary Disease Therapeutics -- Critical Scientific Questions" section does not focus on how best to utilize NIAID's clinical research resources to provide clearer answers on how to use antiretroviral drugs.

In addition, the ad hoc panel which reviewed the ACTG panel criticized the lack of accountability once protocols were underway:

The [Primary Infection] committee's responsibility beyond developing a specific protocol is unclear. There appears to be no systematic accountability for trial logistics. For example, if trials have poor accrual, take too long to complete, have a high drop out rate, or data are not analyzed, specific responsibility appears to be diffuse. Poor co-enrollment, which is scientifically as well as logistically desirable, needs to be addressed because it prevents multi-modality studies from occurring.

One proposal that deserves serious consideration is for all phase 3/4 studies to be largely conducted in CPCRA-type settings, with high priority questions addressed in well-designed, adequately powered studies with minimal data collection, an emphasis on clinical outcomes, comprehensive long-term follow-up, conducted in community primary-care settings. Each of the ACTG's primary infection phase 2/3 studies, with the exception of ACTG 241, is underpowered. Too much data are collected on too few people for too short of a period of time. Unnecessary data points clog the system. ACTG studies continue to focus on laboratory markers, though there has been inadequate validation even for CD4 counts, and there is still no consensus on virological assays. ACTG studies tend to dump participants from studies when they reach their first endpoint, while CPCRA studies follow all clinical events occurring during the study period. Is the time of the first OI really as relevant as the aggregate picture of the intensity of the disease course over a longer period of time? ACTG studies are abominable at carrying out long-term follow-up, especially in early intervention studies. One reason why Concorde's results appeared so compelling was that virtually every participant's fate was recorded. By contrast, in the higher CD4 cohort of ACTG 019, which is now closing out, only 456 of the original 1,620 participants were still being followed earlier this year. The losses to follow-up were so huge (and so biased) that the study may be unreadable.

In order to carry out meaningful phase 3/4 studies, NIAID will have to distribute clinical research resources in a different manner. The CPCRA enrolled one-third as many patients as the ACTG did, in half the time, at one-tenth of the cost. And yet some still feel the CPCRA could conduct simpler, more streamlined studies.

Thus, an emerging paradigm for carrying out studies designed to optimize the antiretroviral standard of care would shift resources from the ACTG model to the CPCRA model. The two networks would collaborate on phase 3/4 studies, with CPCRA enrolling most of the participants, and ACTG carrying out necessary, nested virological, immunological and pharmacological studies. Data would have to be sent to one central site (perhaps the CPCRA statistical unit is more appropriate for this than SDAC). ACTG would continue to carry out phase 1 and 2 (preliminary randomized studies of new agents and new combination regimens, hopefully ones

justified by high-quality preclinical research) studies, and would maintain, possibly even expand, its crucial national network of laboratory-based committees (virology and pharmacology) along with the clinical-laboratory hybrids Immunology, Oncology and Neurology.

Finally, it is critical to demand from those analyzing and disseminating the results of publically-sponsored clinical trials that they tell people the truth about their results. The realities of AIDS are intractable enough without subjecting people to unnecessary joyrides on the rollercoaster of hype and subsequent disappointment. Spurious statistical tools should not be deployed to suggest more clinical benefit than trials actually demonstrate. Particular care should be taken to avoid the interpretive gaffes of Redfield et al in their "responder/nonresponder" analysis, and those of Stanley, Lagakos, and Fischl in their "nontraditional post hoc subgroup trend" analysis from ACTG 155.

The refusal to squarely face the setbacks and surprises which are characteristic of scientific progress in any field, especially AIDS, makes it harder to develop strategies to overcome the setbacks and learn from the surprises.

This week, astronauts on the space shuttle Endeavor are trying to fix the Hubble space telescope, which cannot see the cosmos clearly. It will take more than one week -- and more than just a handful of astronauts -- to fix the mechanism conducting AIDS research. That mechanism, too, cannot see clearly. But it can be fixed, if scientists will start to design studies which can provide clear answers to important questions. It's long past time to turn over antiretroviral clinical trials to a new generation of scientists schooled by AIDS and eager to address the clinical complexities of AIDS competently, openly and honestly.

* * *