Saturday 30 March 2002

BMJ

Initial antiretroviral regimens

In general three drugs are better than two are better than one

Papers p 757

In this issue Rachel Jordan and colleagues have provided a meta-analysis of published controlled clinical trials, which provide significant information relevant to the initial treatment of established HIV infection in adults and adolescents (p 757).[1] Their analyses are based on a large number of well conducted clinical trials. The data largely confirm earlier observations, often from relatively small studies, which have shown that dual nucleoside antiretroviral regimens are significantly more effective than single nucleoside therapy, and that three drug antiretroviral regimens are significantly better than two drug regimens for initial therapy of HIV infection.

The conclusions of the meta-analysis are entirely consistent with current consensus recommendations that initial therapy for established HIV infection in adults and adolescents should include a combination of three potent antiretroviral drugs.[2–4] Since the major biological factor in failure of antiretroviral therapy is the development of viral mutations which confer resistance to specific antiretroviral agents, a two drug regimen has a more durable effect than a single drug regimen, simply because more viral mutations are required to confer resistance to the two antiretroviral drugs. Similarly, resistance to a potent three drug regimen generally takes longer to develop than resistance to a two drug regimen. Clinically relevant resistance may take several years to develop when the patient consistently adheres to a three drug regimen which is successful in decreasing the plasma HIV RNA to less than 50 copies per ml.[5] An additional benefit from a potent three drug regimen appears to be that the viral strains resistant to three antiretroviral agents often exhibit less fitness and virulence than the original wild type virus.

Since antiretroviral naive patients who are adherent to three drug regimens often show effective suppression of HIV RNA levels for several years, it has been difficult to show that a regimen containing four or more drugs is better than a three drug regimen. However, a sound scientific rationale exists for using an initial four drug regimen that includes two protease inhibitors, in which a low dose of ritonavir is used to provide a pharmacokinetic boost to the effectiveness of the second protease inhibitor.

By inhibiting enzymes of the cytochrome P450 system, ritonavir may enhance intestinal absorption (for example, indinavir) or inhibit hepatic metabolism (for example, saquinavir, lopinavir) of certain other protease inhibitors. The low dose ritonavir thus enhances the blood concentration of the second protease inhibitor and thereby decreases the patient's likelihood of developing viral resistance to the second drug. A significant booster effect occurs when ritonavir is given with indinavir, saquinavir, lopinavir, or amprenavir; the effects are greatest with saquinavir, in which a 20 fold increase in the area under the curve may be achieved,[6] and lopinavir, where an up to 100 fold increase in the area under the curve can be obtained.[7]

Other advantages of the coadministration of low dose ritonavir and a second protease inhibitor are that the higher and more prolonged blood concentrations of the second protease inhibitor permit twice daily dosing in all cases. In addition, coadministration eliminates the food restrictions that have been a requirement for giving certain protease inhibitors.

In the case of the combination of ritonavir and saquinavir, the duration of therapeutic saquinavir blood concentrations has been sufficient to justify controlled clinical trials of this combination as a component of once daily, four drug regimens. Data are not yet sufficient to determine long term effectiveness of such regimens.

The meta-analysis presented by Jordan et al provides a solid background against which further changes in therapeutic recommendations can be measured. The meta-analysis, however, deals only with antiretroviral regimens used for initial treatment of established HIV infection in adults and adolescents. It has no direct bearing on the treatment of primary (acute) HIV infection[8] or on the possible use of four or more antiretroviral agents when an initial regimen has failed because of the development of resistance.[9]

Charles Carpenter *professor of medicine*

Department of Medicine, Brown Medical School, Miriam Hospital, 164 Summit Avenue, Providence, RI 02906, USA

1 Jordan R, Gold L, Cummins C, Hyde C. For increasing numbers of drugs in antiretroviral combination therapy. Systematic review and meta-analysis of evidence. *BMJ* 2002;324:757-60.
2 Pozniak A, Gazzard BG, Churchill D, Johnson MA, Williams I, Deutsch JC, et al. British HIV Association (BHIVA) guidelines for the treatment of HIV-infected adults with antiretroviral therapy. BHIVA writing committee on behalf of the BHIVA executive committee. *HIV Medicine* 2001;2:276-313.
3 Carpenter CC, Cooper DA, Fischl MA, Gatell JM, Gazzard BG, Hammer SM, et al. Antiretroviral therapy in adults. Updated recommendations of the International AIDS Society—USA Panel. *JAMA* 2000;283;381-90.
4 Fauci AS, Bartlett JG, Goosby E, Kates J. Guidelines for the use of antiretroviral agents in HIV-infected adults and adolescents. www.hivatis.org/trtgdlns.html, 13 Aug 2001 (accessed 12 December 2001).

BMJ 2002;324:747–8

5 Kaufmann GR, Bloch M, Zaunders JJ, Smith D, Cooper DA. Long-term immunological response in HIV-1 infected subjects receiving potent antiretroviral therapy. *AIDS* 2000;14:959-69.
6 Hsu A, Granneman GR, Cao G, Carothers L, el-Shourbagy T, Baroldi P, et al. Pharmacokinetic interactions between two human immunodeficiency virus protease inhibitors, ritonavir and saquinavir. *Clin Pharmacol Ther* 1998;63:865-70.
7 Sham HL, Kempf DJ, Molla A, Marsh KC, Kumar GN, Chen CM, et. al. aBT-378, a highly potent inhibitor of the human immunodeficiency virus protease. *Antimicrobol Agents Chemo Ther* 1998;42:3218-24.
8 Pantaleo G, Cohen OJ, Schacker T, Vaccarezza M, Graziosi C, Rizzardi GP, et al. Evolutionary pattern of human immunodeficiency virus (HIV) replication and distribution in lymph nodes following primary infection: Implications for antiviral therapy. *Nat Med* 1998;4:341-4.
9 Deeks SG, Hellmann NS, Grant RM, Parkin NT, Petropoulos CJ, Becker M, et al. Novel four-drug salvage treatment regimens after failure of a human immunodeficiency virus type 1 protease inhibitor-containing regimen: Antiviral activity and correlation of baseline phenotypic drug susceptibility with virologic outcome. *J Infect Dis* 1999;179:1375-81.

The long case versus objective structured clinical examinations

The long case is a bit better, if time is equal

The examination of graduates of medicine to ensure competence has a long tradition predicated on the historical right of self regulation bestowed on the professions. While many may wish to replace such summative and frequently punitive assessment with softer assessment to facilitate learning, this amounts to a shirking of social responsibility. A consequence of the importance attached to such examinations is that considerable research has been devoted to establishing the reliability and validity of these examinations.

One truism in educational research is that few self evident truths are true. Historically, it has seemed self evidently true that an experienced physician could, by active questioning around a case, determine whether a candidate was or was not competent—the long case. Unfortunately this assertion was challenged by evidence showing that the reliability of the long case was insufficient to justify decisions about competence to practice.[1] The replacement of the long case by objective structured clinical examinations was predicated on a second self evident truth—the promise of truly objective clinical assessment using checklists, which should self evidently lead to more objective or reliable assessment.

Now, horror of horrors, along comes a study showing that maybe the long case was not so bad after all.[2] This study, published earlier this year in *Medical Education*, involved assessment of 214 final year undergraduates, each of whom did two long cases using patients, and 20 stations for objective structured clinical examinations. After the numbers settled out, the reliability of the long case was 0.84 and of the objective examinations 0.72. The authors conclude that the reliability of long cases is no worse or no better than objective structured clinical examinations in assessing clinical competence.

But perhaps all is not quite what it appears. The long case was observed and evaluated against a "previously used check list which itemised key features of history taking …" as well as global ratings. The quoted reliabilities are for 200 minutes of testing involving 10 long cases or 30 stations for objective examinations. The problem is that no one I know was ever observed doing the long case, no examination ever prepared a detailed scoring sheet in advance, and no candidate ever had an examination with 10 cases. Typically the examination consists of one or two long cases lasting an hour or two where examiners ask their pet questions then give an overall score.

These differences are critical. Performance on one problem is a poor predictor of performance on the next one.[3] So the only solution is to sample many problems. That is the real strength of the objective structured clinical examinations and the real weakness of the traditional long case exam. Still, for equal testing time, the long case turns out a bit better than the objective structured clinical examinations.

What happened to all the gains from the standardisation in the objective structured clinical examinations? Well, maybe they are an illusion after all. Since the variability of performance across problems is the major cause of poor reliability in assessment, efforts to standardise what happens within a case are likely to lead to only small gains.[4] Further, in the long case, examiners used global ratings; in the objective examinations, they exclusively used checklists. And detailed objective checklists turn out to be less reliable than ratings.[5] Perhaps the superiority of the long case in this study is related to using rating scales, not cases which were not standardised.

A companion paper to this study is a further evaluation of the entire examination.[6] This included four written examinations—a multiple choice, true or false test of basic factual knowledge; an extended matching test of problem solving skills; a short answer test of problem solving and data interpretation skills; and an essay to assess ability to present written debate and communicate with professional colleagues. The question was how best to add up the subscores. The answer turned out to accord with best statistical principles. The optimum approach, in terms of maximising reliability, was to weight according to the number of items. But this obscures an underlying philosophical dilemma. Combining scores on subtests which are supposed to measure different things amounts to an admission that they are not so different after all. The fine print in the paper actually confirms this—all the correlations between tests except two or three are in the mid range.

The explanation for this finding is twofold, and is a second example of self evident truths that are not. Firstly, despite the claims of their inventors, different testing formats do not necessarily measure different things.[7] Secondly, the notion that problem solving skills or communication skills can ever be separated from the content of the problem and assessed separately is

BMJ 2002;324:748–9