Reviewer's report

Title: Validity of self-reported male balding patterns

Version: 1 Date: 21 October 2004

Reviewer: Dominique van Neste

Reviewer's report:

General
Accurate hair evaluation remains a difficult task. This is highlighted by the many different global
evaluation techniques proposed since the early publication by Hamilton, and every method has its
drawbacks and some limitations (reviewed in 1).
Patterned hair loss in male and female patients is the end result of a continuum, the ultimate unit
being a recurrent deficiency of hair replacement. A pattern results as the process affects more
severely certain patches generally located on the frontal-parietal areas or on the vertex and some
systems tentatively match the continuum of the phenomenon either in relation with a density scale 2
or a more global coverage evaluation (SCS) 3. The latter SCS system has been calibrated with
refined hair analysis and validated for use in clinical trials and drug evaluation (4, and Van Neste et
al., in press).

One characteristic that is common to all "cartoons" is that no one furnishes further instructions for
the user. The observer (clinician, investigator, patient,...) using these diagrams to the best of his/her
perception, then squeezes the patient into one category. Little is known - and nothing appears to
have been published - about the efficiency or quality of the job of clinical investigators (personal
information). So any effort to document the "value" of categorical classification systems is more than
welcome.

The first point made by the authors is, that choosing one scoring system for patterned scalp loss
implies that those who shall use the system should be trained to use the system, although no
information is being provided about this training and no comparison is given one differences when
inexperienced and experienced or trained observers are compared. Personal evaluation and
observation of private (family) photographs does not provide a critical answer to the question on the
influence of a specific training programme.

The second point that came out clearly from the study is related to "cultural factors". Indeed, age and
educational background appeared as significant factors influencing validity of the personal
evaluation, suggesting that with increasing time balding males become more aware of the exact
degree of balding. This is also an area where educational level may help in a more objective
self-evaluation.

The third aspect is that one has to be satisfied with a 50-75% reliable measurement. Amazingly, the
75% exact duplicate scoring matches the frequency is by chance also recorded in duplicate expert
rating tests (Van Neste et al., in press). Another way of putting things is that one appears deemed to
accept a 25-50% error in the numbers appearing in our records.

The mathematical - statistical consequences of this should be considered as food for future thought.
First, using scores amongst the inclusion criteria for a given study (clinical trial, epidemiological
evaluation, etc...) is not a major problem if one accepts that untrained observers and self-rating,
especially in the younger and less-educated population, may accept up to 50% of subjects that do
not match the inclusion criteria.
Second, as far as drug trials are concerned, use placebo controls. There is no evidence that "improvement" in treated subjects or "worsening" in placebo treated controls will move from one category to another in the time lapse of the trial. This raises a recurrent ethical consideration: why should rough estimators of disease i.e. imprecise and inappropriate methods continue to be promoted for use in clinical trials as such a choice imposes placebo treatment for a chronic progressive disorder? Third, is the grouping categories or averaging scores a reasonable thing to do? In a sub study, grouping of 4 different categories into one, improved the correlation from 67% to 87 %. This is no surprise as a one or even two category errors or difference are simply wiped out! If all severity classes were to be grouped into one category i.e. male pattern baldness (all severities) one would predict a 100% agreement between ratings. How could one evaluate that 87% is significantly different from the 75% exact agreement, apparently a maximum level reached only by properly trained experts!

Finally, "time" is an important dimension that was not considered in this assay. Indeed, in essence, patterned hair loss is dynamic process. As this is a retrospective study and that family photographs may not be considered as standardized documents, one may hope that a more standardized approach for global scalp imaging may be considered as a nice project for a future trial

references
in press:
Van Neste D, Sandraps E, Herbaud D, Lelubre P, Leroy T: Validation of scalp coverage scoring methods for scalp hair loss in male pattern hair loss (androgenetic alopecia). Skin Research and Technology, accepted for publication

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

Discretionary Revisions (which the author can choose to ignore)