Ann Hematol (2001) 80:B58–B60

Dirk Hasenclever · Markus Loeffler

Five theses concerning the clinical consequences of pathology and prognostic factors

Abstract We discuss possible justifications to split study populations from a biometrical point of view. The existence of prognostic differences between subgroups are neither a sufficient nor a necessary reason to justify a splitting decision. There are essentially two separate types of relevant arguments to justify a split of patient study populations: a) Different toxicity/benefit trade-offs concerning the acceptability of a particularly aggressive treatment, b) Evidence for strong treatment by subgroup interactions, i.e. Treatment differences differ markedly by biologically defined subgroups. The latter is what the research ideal of biologically specific treatment asks for. Subgroup analysis is notoriously difficult. Formal statistical analysis must be complemented by specific evidence from basic sciences. Meta-subgroup analyses may be an option if a biologically specific hypothesis on which treatment component interacts with what biological feature allows to operationally identify all those randomised trials in which the effect should be present.

In this conceptual and methodological paper we defend five theses concerning the clinical consequences of pathological and biological differences from a somewhat unorthodox biometrical point of view.

1. Splitting study populations

Thesis 1:

Rash fragmentation of study populations into differently treated groups undermines the epistemological basis to investigate and understand the specific clinical impact of new pathological and biological differences.

Studies to determine the clinical relevance of biological features have the objective to demonstrate that patients with specific parameters or subentities respond differently to well defined treatment strategies. To demonstrate this, the patients in question have to be treated within a common trial. This advocates for broad inclusion spectra. In addition, study population fragmentation tends to promote statistically under-powered studies.

Splitting study populations must therefor be considered with great reservation and should be rigorously justified.

D. Hasenclever (✉)
University of Leipzig, IMISE, Liebigstr. 27,
D-04103 Leipzig, Germany
e-mail: Hasenclever@IMISE.uni-Leipzig.de

2. The relevance of prognostic factors

Thesis 2:

The question whether a new pathological entity or biological marker carries prognostic information is of minor relevance when discussing whether to split study populations.

Analyses of prognostic factors tacitly assume that prognostic differences are independent of treatment differences as long as ‘state of the art’ treatment strategies are employed (Note that we are here particularly referring to treatments like polychemotherapy). Prognostic indices typically are derived from large data bases pooling data from several study groups since they require substantial patient numbers. By pooling results from different treatment schemes and different inclusion spectra the resulting indices apply to an ‘average’ treatment disregarding potential differences in treatment outcomes. This averaging-out is generally intended in prognostic factor analyses and the result is interpreted to be a general biological feature of the disease (heterogeneity) rather than the treatment. Thereby prognostic factors provide insight into the disease biology and eo ipso help to judge prognosis.

The very fact that a biological difference is prognostic, however, does not by itself suggest that the respective patient subgroups should be treated differently. On the contrary, consistent with the underlying intention of prognostic factor analyses that prognostic factors are independent from (minor) treatment differences, a more

successful treatment in one subgroup should be expected to improve results also in the other subgroup. Thus prognostic information is not a sufficient criterion to split a study population.

One the other hand, it is conceivable that a group of patients turns out to consist of two quite distinct entities which have the same prognosis with unspecific treatment. One of them may, however, be accessible to a specific (e.g. antibody) treatment against a biological locus. Thus prognostic information is not a necessary information to split a study population.

Prognostic factors are neither a sufficient nor a necessary reason to justify a splitting decision.

The improvement of prognostic scores by integrating biological markers or pathological differences on the one hand and the search for clinical consequences of pathological and biological differences on the other are only loosely related areas of research and should not be confused.

3. Arguments to justify splits in patient populations

Thesis 3:

There are essentially two separate types of relevant arguments to justify a split of patient study populations:

a) Different toxicity/benefit trade-offs concerning the acceptability of a particularly aggressive treatment.

When comparing a standard treatment with an aggressive experimental approach one has to weigh the potential benefit in disease control versus the higher toxicity burden of the experimental arm in defining the appropriate study population. The danger is that toxicity will dilute the possible benefit in disease control. Thus one would restrict the study to patients in which toxicity is expected to be manageable (possibly based on prognostic factors for toxicity, e.g. age and performance status etc.) and/or which have such a dismal prognosis (based on prognostic scores for treatment outcome) that high toxicity rates appear to be acceptable.

If – for illustration – the aggressive treatment is expected to have a 10% acute mortality rate and the treatment associated mortality with standard treatment is negligible, a patient with 90% chance of cure cannot possibly benefit from the more aggressive treatment. On the other hand, a high risk patient with about 40% chance of cure with standard treatment might profit if the gross difference in lymphoma control is appreciably greater than the product of (excess mortality rate)×(cure rate with standard treatment), i.e. 4% in this example. Trials of high dose chemotherapy versus conventional treatment often are restricted to younger high risk patients due to such toxicity/benefit consideration.

To make this type of argument explicit, one needs prognostic scores for general treatment outcome and prognostic information on toxicity disposition and a scenario of both expected toxicity levels and treatment outcome with both the standard and the experimental treatment.

This type of justification is pragmatic, ad hoc and depends only in so far on biological factors as they have become integrated into prognostic scores for outcome or toxicity.

b) Evidence for strong treatment by subgroup interactions:

A treatment by subgroup interaction is present if the outcome differences between two treatments are not similar across all subgroups but differ between them in a systematic way. A strong treatment by subgroup interaction (i.e. difference of differences) is a sufficient reason to split study populations. We claim that this criterion is of prime relevance in any discussion of clinical consequences and has so far largely been ignored.

If – for example – addition of B-cell antibody treatment would prove generally beneficial in B-cell disease (in several trials with varying inclusion criteria and CT-treatments), this specific treatment option will justify to split T-cell from B-cell lymphoma study populations. Optimal combinations of antibody and CT-treatment will then be developed in the B-cell population.

There may also be less but still sufficiently specific interactions: There is same evidence that very dose intense induction plus high dose chemotherapy with stem cell support may be beneficial in rapidly growing high risk lymphoma while not improving results in low risk cases (Compare Hasenclever et. al. in this volume for discussion and interpretation). If this finding were confirmed in several trials a split of low vs. high risk populations would be justified.

This type of justification is much more basic and specific than the toxicity based argument: The detection and exploitation of strong treatment by subgroup interactions is a central prerequisite to develop biologically specific treatment.

4. Searching for treatment by subgroup interactions

One has to do subgroup analyses to search for clinically relevant treatment by subgroup interactions. Results of subgroup analyses are notoriously difficult to interpret because of an underlying fundamental combinatorial problem:

A strong treatment by subgroup interaction is observed when the efficacy of a particular treatment component depends on the presence of a specific biological feature (that defines the subgroups). But there is a plethora of biological differences on the disease side and a huge number of treatment parameters (drugs, doses, timing etc) on the therapeutic side. The number

of conceivable combinations is even larger and almost indefinite.

Thesis 4:

Based on analysis of clinical data alone it is (nearly) hopeless to identify which treatment characteristic interacts with which biological feature. But just that is required for intelligent biologically specific treatment.

In addition to this basic combinatorial problem, subgroup analyses are complicated by two other more statistical problems:

1. Looking at a huge number of statistical hypotheses inflates the danger of spurious false positive findings (multiplicity problem) and
2. Data from single clinical trials tend to be statistically under-powered to do treatment by subgroup interaction analyses since there size is usually determined assuming a uniform treatment effect across subgroups. Trial sizes required for subgroup analysis may be substantially larger.

Statisticians are therefor generally extremely reluctant to do subgroup analyses. The formal statistical problem probably has no satisfactory solution, except perhaps huge unrealistically large patient numbers.
Side remark: Several statistical 'dogmas' are upheld as remedies in order to safeguard against misleading spurious findings:

"Never conduct a subgroup analysis if there is no overall treatment effect!" is one of them. But it is easily conceivable that a treatment component has a strong effect restricted to a limited subgroup, such that overall the treatment effect is not "significant" due to limited statistical power.

Often formal pre-specification of the interaction hypotheses in the study protocol is required before doing a subgroup analysis to counteract the multiplicity problem. But again, this requirement is implausible if the accumulating data of study groups all over the world is regarded as a common resource for the scientific community. From this "meta" point of view, the difference between an originally planned and a secondary analysis of the data is blurred.

These type of preventive remedies perhaps insure against spurious false positive findings, but tend to be extremely conservative and clearly do not help with the identification problem. They are justified in situations where interactions are generally implausible, but may be counterproductive, when biological specificity is a research focus.

How to establish a treatment by subgroup interaction

Thesis 5:

Formal statistical analysis must be complemented by evidence from basic sciences to overcome or avoid these problems (particularly the combinatorial one):

What is needed in order to establish a treatment by subgroup interaction is

a) first a specific interaction hypothesis,
1) naming a specific biological feature (defining the subgroups)
2) naming a particular treatment component or principle the efficacy of which is proposed to depend on the presence of the biological feature
3) supporting the link by some specific basic evidence or theory.

b) then, given such a specific interaction hypothesis, one can clearly identify in which (type of) randomised trials this interaction should be expected to be observable. Thus provides the inclusion criteria (on a trial basis) for a targeted meta-subgroup analysis, that eventually provides the required clinical evidence.

Specificity of interaction hypothesis arguable is a matter of degree to a certain extent. But the decisive operational point is the following: With a specific interaction hypothesis it is determined in which randomised trials one would expect to be able to find the interaction if it existed. The interaction is specific and item b) is indicated if and only if the hypothesis operationally defines inclusion criteria for a confirmative meta-subgroup analysis. For unspecific interaction hypotheses it is indefinite where to look for confirmatory replication.

Results from explorative subgroup analyses should be carefully reported, but should be regarded as 'curious findings' by the scientific community, unless and until a qualified interaction hypothesis explaining them is presented and the interaction is independently replicated or confirmed in a meta-subgroup analysis.