What Do We Mean by ‘Replication’ and ‘Validation’ in Genome-Wide Association Studies?

Bernd-Wolfgang Igl  Inke R. König  Andreas Ziegler
Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Lübeck, Germany

Key Words
Genome-wide association · Replication · Validation

Abstract
It has been acknowledged that results from genome-wide association studies need to be reproduced in further independent samples. This includes both aspects of replication and validation. Although often used interchangeably, both of these terms have a specific unique meaning and are not synonyms. Based on a number of applications from the literature producing a certain amount of confusion, our aim therefore is to clarify the definition of ‘replication’ and ‘validation’ and propose standards of their usage.

Despite the agreement on the necessity of confirmation, there seems to be some confusion regarding the terms used for this endeavor. Specifically, confirmation studies often aim at ‘replicating’ or ‘validating’ previously described associations without defining what this attempt comprises. For example, Greenwood et al. [3] use the terms ‘validation’ and ‘replication’ interchangeably to describe that ‘two or more independent datasets demonstrate the same disease-marker associations, when analyzed independently’ (p. 397). The same use of terms can be found in applied works on GWAs [4]. Other groups employ exclusively either the term ‘validation’ [5] or ‘replication’ [6, 7].

The different use in terminology is complicated by the fact that the similarity between original and confirmation samples varies to a great extent between studies. For example, the samples may be from different geographic regions (e.g., Iceland, Sweden and Hong Kong [6], and United Kingdom and Germany [7]), or different sampling strategies and different genotyping technologies may have been used [5].

For a long time, it has been acknowledged that results from genetic association studies need to be confirmed in further independent samples. Reasons for this are manifold and include the desire to reduce systematic bias caused, for instance, by population stratification [1]. In the realm of genome-wide association studies (GWAs), the need for confirmation has become imperative. Here, in addition to systematic errors, screening hundreds of thousands of markers for association has dramatically increased the chance for false positives based on the multiple testing problem [2].

B.-W. Igl and I.R. König contributed equally to the work presented in this paper.
differences. This, in turn, makes the interpretation of results difficult, especially if the confirmation fails. Consider, for example, the case of a GWA conducted with central European probands, the results of which cannot be confirmed in a Japanese sample. With the current practice of use of terminology, ‘replication has failed’, but this does not express that the cause may be systematic population differences instead of random variation.

Given these complications, we consider it intuitively to reserve different terms for the confirmation attempt with different degrees of variation between populations. Specifically, we suggest distinguishing two scenarios (see fig. 1):

(1) ‘Replication:’ (Latin: replicatio = a folding or rolling back again, a reply), both original and confirmation sample are drawn from the same population, and systematic differences are reduced to a minimum. Thus, differences in results are largely due to random variation so that results can be confirmed in dependence of sample size and the amount of randomness within the study. Following the proposal by Chanock et al. [8], this design aims at a ‘replication’ of the GWA. In detail, they defined specific criteria for establishing replication, including that a similar population and the same study design and analysis must be used. Similarly, Clarke et al. [9] described as gold standard for ‘replication’ the repeated association under identical circumstances using an independent sample. In this sense, a result can be ‘replicated’ with the aid of an internal cross validation.

(2) ‘Validation:’ (Latin: validus = mighty, powerful, robust, strong), the confirmation sample stems from a population which is different than that from which the original sample was drawn. Differences between these populations may concern the ethnic background, the phenotype definition, the recruitment or sampling strategy, and the time point of investigation. As a result, differences in the outcome may be explained both by random and systematic variability. Using this kind of design, ‘validation’ of the GWA is attempted. More precisely, to ‘validate’ a result means to obtain similar findings under modified influencing factors such as ethnicity, phenotype, time, etc., for example using an external cross validation design.

We focus on these two terms since they are widespread and common used in different statistical applications [10–14]. These definitions are independent of a retrospective or prospective design and yield consequences regarding the interpretation of confirmation studies in general. If a genetic association has been ‘validated’, it shows greater generalizability than a ‘replicated’ association.

With this definition, the terms replication and validation are similar to the terms temporal validation and external validation which are used in clinical epidemiology [15].

It should be noted that these specifications are our suggestions and should be a basis for further discussions. In general, our aim was to clarify the use of terms and consequently, we did not take strict definitions of additional vocabulary like ‘confirmation’, etc. into account.

References


