Let's call it the effect allele: a suggestion for GWAS naming conventions

International Journal of Epidemiology, 2020, 1734–1735
doi: 10.1093/ije/dyaa149
Advance Access Publication Date: 2 September 2020

Robyn E Wootton[1,2]* and Hannah M Sallis[1,2,3]

[1]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK, [2]Department of Population Health Sciences, Bristol Medical School, University of Bristol, UK and [3]School of Psychological Science, University of Bristol, Bristol, UK

*Corresponding author. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. E-mail: robyn.wootton@bristol.ac.uk

In recent years, the amount of publicly available summary data from genome-wide association studies (GWAS) has rapidly increased. So too has the number of researchers accessing and utilizing these data. These summary data can be used in many subsequent analyses, including Mendelian randomization, genetic correlation and polygenic score analysis. It is therefore vital that we can ensure consistency across these datasets to minimize the risk of analytical mistakes due to user error. One such inconsistency is the naming of the effect allele in these datasets. This is of particular concern given the increasing availability of automated software packages for downstream analyses (e.g. MR-Base,[1] LDpred,[2] LD Hub[3]) which make complex analyses easy to run, even with little understanding of the underlying processes. Data harmonization is a crucial step for many of these analyses. Guidelines for harmonizing datasets to use in two-sample Mendelian randomization have been suggested by Hartwig *et al.*[4] In order to effectively implement these steps, it is critical to know which is the effect allele. This starting point is assumed to be known in the Hartwig *et al.* paper, however, in many datasets this is unclear and could lead to substantial errors. This issue could be easily avoided if consistent naming conventions were used.

We have seen them all: A1/A2, A0/A1, effect allele/non-effect allele, effect allele/other allele, reference (REF) allele/alternative (ALT) allele, the list goes on. This difference in terminology can be due to the software used to perform the GWAS [e.g. plink:[5] ALT (effect)/REF (other); SNPTEST:[6] A(other)/B (effect)] or simply due to the naming convention used by the analyst when they create the datasets. If these terms were used consistently it would only be a minor annoyance to learn the different naming conventions. However, in GWAS summary datasets the terms A1/A0/alternative allele do not consistently refer to the effect allele [e.g. Liu *et al.*:[7] ALT (effect)/ REF (other); Snieckers *et al.*:[8] REF(effect)/ALT (other)]. Sometimes they refer to the non-effect allele, or even to the minor allele (the variant that is less frequently found in the population). The effect allele is the allele to which the effect estimate refers, regardless of whether this estimate is increasing or decreasing and regardless of whether this allele is coding or non-coding. Sometimes, the minor allele is used as the effect allele but the two are not synonymous. The major allele (most frequently found in the population) can also be chosen as the effect allele. Therefore, knowing the minor allele alone is not sufficient for the majority of downstream analyses.

In order to work out which is the effect allele, researchers are reliant on clear documentation from the authors of the GWAS—this is often not available or is lacking in detail. When it does exist, it can be hard to find and is often buried deep in a supplementary table. We have experienced much confusion and many misunderstandings that could easily be avoided if GWAS authors were to use the terms 'effect allele' and 'other allele' instead. Despite the publication of reporting guidelines for genetic association studies [STrengthening the REporting of Genetic Association Studies (STREGA)[9]] there is currently no naming convention for reporting the effect allele. We would encourage this as an addition to point 16 (Main Results) of the current STREGA guidelines.

We are not trying to say that introducing this naming convention would be a panacea, and there are many other pieces of information and metadata that would help with the harmonization process, as discussed elsewhere.[10] For example, information on effect allele frequency or strand alignment and genomic build would help with the harmonization of ambiguous variants (e.g. palindromic variants with intermediate allele frequency). However, we believe that introducing a consistent allele naming convention is a simple suggestion that would be straightforward to implement. This simple step could go a long way to making data harmonization more efficient and less error prone.

Funding

This work was supported by the Medical Research Council and the University of Bristol (MC_UU_00011/1, MC_UU_00011/7). H.M.S. is also supported by the European Research Council (Grant ref: 758813 MHINT).

References

1. Hemani G, Zheng J, Elsworth B *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *eLife* 2018;**7**:e34408.
2. Vilhjálmsson BJ, Yang J, Finucane HK *et al.* Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet* 2015;**97**:576–92.
3. Zheng J, Erzurumluoglu AM, Elsworth BL *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* 2017;**33**:272–9.
4. Hartwig FP, Davies NM, Hemani G, Davey Smith G. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. *Int J Epidemiol* 2016;**45**:1717–26.

5. Purcell S, Neale B, Todd-Brown K *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;**81**:559–75.
6. Marchini J, Band G. *SNPTEST v2 Technical Details*; 2010.
7. Liu M, Jiang Y, Wedow R *et al*. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* 2019;**51**:237–44.
8. Sniekers S, Stringer S, Watanabe K *et al*. Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat Genet* 2017;**49**:1107–12.
9. Little J, Higgins JP, Ioannidis JP *et al*. STrengthening the REporting of Genetic Association Studies (STREGA)—an extension of the STROBE statement. *Genet Epidemiol* 2009;**33**:581–98.
10. Lyon MS, Andrews SJ, Elsworth BL *et al*. The variant call format provides efficient and robust storage of GWAS summary statistics. *bioRxiv* 2020.05.29.115824; 2020. doi: 10.1101/2020.05.29.115824.

BCG vaccination in Bangladesh: should it be given at birth or given along with pentavalent?

International Journal of Epidemiology, 2020, 1735–1738
doi: 10.1093/ije/dyaa172
Advance Access Publication Date: 18 September 2020

Syed Manzoor Ahmed Hanifi[1]* and Mizanur Rahman[2]

[1]Health Systems and Population Studies Division, International Centre for Diarrhoeal Disease Research, Bangladesh and [2]Carolina Population Center, University of North Carolina at Chapel Hill, NC, USA

*Corresponding author. Health Systems and Population Studies Division, International Centre for Diarrhoeal Disease Research, Bangladesh. E-mail: hanifi@icddrb.org

Bangladesh is a tuberculosis (TB) endemic area and the TB incidence rate has remained the same for over a decade.[1] Most of the exposed become infected within a year of initial exposure[2] and the disease progresses fastest in neonates.[3] This is why early BCG vaccination is imperative. The World Health Organization (WHO) recommends BCG vaccination at birth followed by the combined diphtheria, tetanus and whole-cell pertussis (DTP) vaccine at 6, 10 and 14 weeks (we term it Schedule A in Table 2), but in practice the Bangladesh Extended Program on Immunization (EPI) by and large provides BCG and pentavalent together at 6 weeks of age or later (we term it Schedule B in Table 2).

The Bangladesh EPI has achieved almost universal child vaccination coverage through its monthly outreach sessions along with fixed facility services, but a few newborns receive BCG at birth, and only 20% receive it by the first month.[4] When EPI's extensive outreach programme was developed in the 1980s most births took place in homes (<5% of deliveries took place at facilities),[5] so it was not feasible to provide BCG at birth or within a few days of birth through a door-step approach. Moreover, due to some traditional beliefs mothers had a preference for not taking their newborns out of home in the early days of life, but this is changing with increased access to healthcare in fixed-site facilities. In recent years facility-based delivery has increased phenomenally in Bangladesh from just 4% in 1994[5] to 50% in 2017.[6] This rise in facility delivery has opened a window of opportunity to raise coverage of BCG vaccination at birth.

In Bangladesh, one concern against providing BCG at birth was the high prevalence of low birthweight as it was suggested that low birthweight babies should be given BCG at a later age to avoid the vaccine's potential adverse effects. In recent studies Higgins *et al.* 2016,[7] Aaby *et al.*[8] and Biering-Sørensen *et al.* [9] demonstrated that newborns with low-birthweight benefit from the BCG given at birth.

Immunological studies have shown that BCG given to children early in life increases adaptive[10] response to other vaccines.[11] Furthermore studies conducted in murine models suggest that BCG is protective against sepsis and other bacterial, viral and fungal infections.[12] It has also recently been found that the vaccine may also enhance the non-specific innate resistance through epigenetic reprogramming of monocytes.[13] Thus, the vaccine has a broader protective action than just the prevention of targeted disease. A review by WHO's Strategic Advisory group of Experts on Immunization (SAGE) on non-specific effects of BCG vaccine shows that BCG vaccination at birth reduces all cause neonatal mortality among neonates by 48% (18–67%).[7]

In 2016, a study conducted in Spain showed that BCG vaccination at birth is associated with decreased hospitalization due to lower respiratory infection and sepsis among children <15 years of age.[14] WHO recommends that mothers initiate breastfeeding within 1 h of birth and early breastfeeding initiation decreases the risk of neonatal mortality.[15] Several studies state that breastfed infants and preterm birth infants should be vaccinated according to the routine recommended schedules.[16,17] In one study, breastfeeding improved the cellular immune response to BCG vaccine if the vaccine was given at birth. However, if the vaccine was given at 1 month of age, breastfeeding had no effect.[18]

Recently Aaby *et al.*[19] compared the mortality of infants who received BCG alone before DTP with those who received BCG at 6 weeks or later along with DTP, in Matlab, Bangladesh during the period 1986–99. (It appears that the former group resembles Schedule A and the latter Schedule B.) Mortality during ~6 weeks to 9 months was 50% lower among those who received BCG and DTP together (16 per 1000 person-years) compared with those who