

# Two SNPs Associated With Spontaneous Dizygotic Twinning: Effect Sizes and How We Communicate Them

Hamdi Mbarek,<sup>1,2</sup> Conor V. Dolan,<sup>1</sup> and Dorret I. Boomsma<sup>1,2</sup>

<sup>1</sup>Department of Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands

<sup>2</sup>Avera Institute for Human Genetics, Sioux Falls, South Dakota, USA

In a recent GWAS of spontaneous dizygotic twinning, Mbarek et al. (*The American Journal of Human Genetics*, 2016, Vol 98, pp. 898–908) identified two SNPs, rs11031006 (near *FSHB*) and rs17293443 (in *SMAD3*). In the present note, we address the question how to present the results in terms of effect sizes in a manner that is comprehensible to the general audience (e.g., mothers of twins, readership of newspapers). We propose to avoid the standard effect sizes such as odds ratios and relative risk as these require some knowledge of probability theory. Rather, we convey the results in terms of the conditional probabilities, but expressed in natural language.

■ **Keywords:** DZ twinning, genetic variants, effect size

We have brought together unique collections of mothers of dizygotic (DZ) twins, and conducted the first genome-wide association study (GWAS) for spontaneous DZ twinning (Mbarek et al., 2016). We identified the first robust genetic risk variants for DZ twinning: one near *FSHB* (rs11031006; chr 11p14.1), and a second within *SMAD3* (rs17293443; chr 15q22.23). The two signals, which were replicated in a large Icelandic cohort, turned out to have widespread effects on female fertility. In disseminating the results of GWAS studies, like Mbarek et al. (2016), one tends to emphasize the genome-wide significant hits. However, communicating information on the effect sizes associated with these hits to the general public is not simple. Standard effect sizes, such as odds ratios (OR) or relative risks (RR), are largely unsuited as they require some familiarity with probability theory. However, to rise above the simple ‘gene for...’ statement, it is desirable to provide a comprehensible statement on effect. The aims of this note are: (1) to present the results of the Mbarek et al. (2016) study in terms of conditional probabilities, and (2) to consider, in the light of these, various effect sizes based on these results, given the aim of communicating the results to the general population.

## Derived Genotype — Twinning Probabilities

To ease presentation, we denote the SNPs FB (rs11031006) and S3 (rs17293443). Based on Mbarek et al. (2016), the risk

allele frequencies (RAFs) in Iceland are 0.85 (FB; risk allele G, alternative allele A) and 0.24 (S3; risk allele C, alternative allele T). We use these values as they coincide with meta-analytic estimates, and there is very little difference in RAF between the samples in Mbarek et al. (2016) (Netherlands: 0.841 and 0.239; Australia: 0.863 and 0.239; and United States, Minnesota: 0.846 and 0.233). The meta-analytic values are 0.849 and 0.236). The RR associated with these risk alleles, given the additive genetic association model, are 1.18 (FB) and 1.09 (S3). These results were obtained in the Icelandic sample (i.e., the replication sample in Mbarek et al., 2016), and we assume here that they hold in the Dutch population. The prevalence of spontaneous DZ twinning in the Dutch population displays considerable variation between 1904 and 2011 (Glasner et al., 2013; see also Pison et al., 2015). We used the most recent prevalence estimate reported by Glasner et al. (2013), that is, ~1.07% in 2011. The loci FB and S3 are unlinked (located on chromosomes 11 and 15), and, we assumed, in gametic phase equilibrium (Wray and Visscher, 2007). From this information and the assumptions mentioned, we constructed

RECEIVED 4 May 2016; ACCEPTED 10 May 2016. First published online 13 July 2016.

ADDRESS FOR CORRESPONDENCE: Hamdi Mbarek, Vrije Universiteit, Dept Biological Psychology, Netherlands Twin Register, Van der Boechorststraat 1, 1081 BT, Amsterdam, the Netherlands. E-mail: [h.mbarek@vu.nl](mailto:h.mbarek@vu.nl)

**Table 1.**

FSHB (FB) and SMAD3 (S3) genotypes by spontaneous twinning probabilities. Row 9 and column 5 contain the marginal probabilities (*italics*; .0107 is the 2011 Dutch spontaneous DZ twinning probability). As indicated, the genotype frequencies are a function of the risk allele frequency (denoted  $p_{FB}$  and  $p_{S3}$ ;  $q_{FB}=1-p_{FB}$  and  $q_{S3}=1-p_{S3}$ ).

genotype	N risk alleles	DZ Yes	DZ No	genotype frequencies
FB(G;G)	2	0.0081	0.7144	<i>0.7225</i> ( $p_{FB}^2$ )
FB(A;G)	1	0.0024	0.2526	<i>0.2550</i> ( $2*p_{FB}*q_{FB}$ )
FB(A;A)	0	0.00018	0.0223	<i>0.0225</i> ( $q_{FB}^2$ )
		<i>0.0107</i>	<i>0.9893</i>	
S3(C;C)	2	0.0007	0.0569	<i>0.0576</i> ( $p_{S3}^2$ )
S3(C;T)	1	0.0041	0.3607	<i>0.3648</i> ( $2*p_{S3}*q_{S3}$ )
S3(T;T)	0	0.0059	0.5717	<i>0.5776</i> ( $q_{S3}^2$ )
		<i>0.0107</i>	<i>0.9893</i>	<i>1</i>

**Table 2.**

FSHB (FB) and SMAD3 (S3) haplotypes by spontaneous twinning probabilities. Row 11 and column 5 contain the marginal probabilities (*italics*; note: .0107 is the 2011 Dutch spontaneous DZ twinning probability). As indicated, the genotype frequencies are a function of the risk allele frequencies (denoted  $p_{FB}$  and  $p_{S3}$ ;  $q_{FB}=1-p_{FB}$  and  $q_{S3}=1-p_{S3}$ ).

haplotype	N risk alleles	DZ Yes	DZ No	haplotype frequency
FB(G;G) & S3(C;C)	4	0.00053	0.04109	<i>0.04162</i> ( $p_{FB}^2*p_{S3}^2$ )
FB(G;G) & S3(C;T)	3	0.00302	0.26054	<i>0.26357</i> ( $p_{FB}^2*2*p_{S3}*q_{S3}$ )
FB(G;G) & S3(T;T)	2	0.00453	0.41279	<i>0.41732</i> ( $p_{FB}^2*q_{S3}^2$ )
FB(A;G) & S3(C;C)	3	0.00016	0.01453	<i>0.01469</i> ( $2*p_{FB}*q_{FB}*p_{S3}^2$ )
FB(A;G) & S3(C;T)	2	0.00093	0.09209	<i>0.09302</i> ( $2*p_{FB}*q_{FB}*2*p_{S3}*q_{S3}$ )
FB(A;G) & S3(T;T)	1	0.00135	0.14594	<i>0.14729</i> ( $2*p_{FB}*q_{FB}*q_{S3}^2$ )
FB(A;A) & S3(C;C)	2	0.000012	0.00128	<i>0.00130</i> ( $q_{FB}^2*p_{S3}^2$ )
FB(A;A) & S3(C;T)	1	0.000070	0.00814	<i>0.00821</i> ( $q_{FB}^2*2*p_{S3}*q_{S3}$ )
FB(A;A) & S3(T;T)	0	0.000101	0.01289	<i>0.01299</i> ( $q_{FB}^2*q_{S3}^2$ )
		<i>0.0107</i>	<i>0.9893</i>	<i>1</i>

**TABLE 3**

Conditional Probabilities Prob (DZ|genotype) and RRs and OR Associated with FB (Reference A;A) and S3 Genotypes (Reference T;T)

Genotype	N risk alleles	RR	OR	Prob(DZ genotype)
FB(G;G)	2	1.392	1.397	0.0112
FB(A;G)	1	1.180	1.182	0.00950
FB(A;A)	0	1.000	1.000	0.00805
S3(C;C)	2	1.188	1.190	0.01218
S3(C;T)	1	1.090	1.091	0.01117
S3(T;T)	0	1.000	1.000	0.01025

Note: These results are based on the probabilities given in Table 1.

the genotype by twinning probability tables (Tables 1 and 2; for details, see the Appendix; R scripts used as available as supplemental material).

### Effect Sizes: Odds Ratios and Relative Risks

ORs and RRs are standard measures of effect size, given a binary outcome. These measures are shown in Table 3 (separately for FB and S3) and in Table 4 (haplotypes). The values were calculated relative to the zero risk allele genotypes and haplotype. Note that the ORs and RRs are similar because

the prevalence of the outcome is low (i.e., 0.0107). We have also included the probabilities of a spontaneous DZ birth given (conditional on) the genotype (Table 3) and haplotype (Table 4), as we propose to focus on these probabilities in communicating the effects.

### How to Convey These Results to a General Audience

Effect sizes expressed in terms of ORs and RRs pose no problem for (genetic) epidemiologists. However, as their interpretation requires knowledge of probability theory, we consider them unsuited for the general readership of newspapers and other media. To convey these effect sizes in simple terms, it is desirable to avoid terms like 'odds', 'odd ratio', 'conditional probabilities', and 'RR'. We propose to present the effects for the combined effects of FB and S3 as follows: 'In the Dutch population the probability of spontaneous DZ twinning is 10.7 per 1,000 births. If the risk alleles were absent, this would be 7.76 per 1,000 births. If all females carried all four risk alleles, this would be 12.71 per 1,000 births.' This is an attempt to express, in natural language, the effect size in terms of the conditional probabilities associated with the double homozygotes (FB(A;A) & S3(T;T)

**TABLE 4**  
Allelic ORs and RRs Associated With the Two SNPs

Haplotype	N risk alleles	RR	OR	Prob(DZ haplotype)
FB(G;G) and S3(C;C)	4	1.637	1.645	0.01271
FB(G;G) and S3(C;T)	3	1.478	1.483	0.01147
FB(G;G) and S3(T;T)	2	1.397	1.401	0.01085
FB(A;G) and S3(C;C)	3	1.397	1.401	0.01085
FB(A;G) and S3(C;T)	2	1.289	1.292	0.01001
FB(A;G) and S3(T;T)	1	1.179	1.181	0.00916
FB(A;A) and S3(C;C)	2	1.182	1.184	0.00918
FB(A;A) and S3(C;T)	1	1.092	1.093	0.00848
FB(A;A) and S3(T;T)	0	1.000	1.000	0.00776

Note: The ORs and RRs associated with the haplotypes (reference A;A and T;T). These results are based on the probabilities given in Table 2.

and FB(G;G) and S3(C;C)). Note that the value 10.7 is based on the 2011 Dutch spontaneous twinning probability (i.e., 0.0107; Glasner et al., 2013), and the values 7.76 and 12.71 are  $\text{prob}(DZ|\{FB(A;A) \& S3(T;T)\})$  and  $\text{prob}(DZ|\{FB(G;G) \& S3(C;C)\})$ , respectively (see Table 4, column 5). Similar statements can be formulated for FB and S3 in isolation.

We believe that the expression (in natural language) of effect size relating to a continuous phenotype is relatively simple, as one can avoid statistical terms like ‘variance explained’ and ‘correlation,’ by relating the allelic effect directly to the scale used to measure the phenotype. However, such statements of effect size remain natural language expressions concerning conditional distributions (rather than conditional probabilities). For instance, Loos and Yeo (2014), in their discussion of the effects of the FTO locus, related the risk allele to a 0.39 kg/m<sup>2</sup> increase in body mass index and to an increase in weight of 1,130 g for a person of 1.70 m.

DZ twinning is obviously a polygenic trait. Polygenic risk scores results in Mbarek et al. (2016) reflect the polygenic contribution to the susceptibility to DZ twinning and its association with greater reproductive ability. Revealing more signals associated with this trait will provide a clear picture of the risk prediction for having DZ twins.

## Acknowledgments

Support for the Netherlands Twin Register was obtained from the Netherlands Organization for Scientific Research (NWO) and The Netherlands Organization for Health Research and Development (ZonMW) grants, 904-61-193,480-04-004, 400-05-717, Addiction-31160008, 911-09-032, Biobanking and Biomolecular Resources Research Infrastructure (BBMRI -NL, 184.021.007); Royal Netherlands Academy of Science Professor Award (PAH/6635) to DIB; European Research Council (ERC-230374 and ERC-284167); Rutgers University Cell and DNA Repository (NIMH U24 MH068457-06), the Avera Institute, Sioux Falls, South Dakota (USA) and the National Institutes of Health (NIH R01 HD042157-01A1). Part of the genotyping was funded by the Genetic Association Information Network (GAIN) of the Foundation for the National

Institutes of Health and Grand Opportunity grants 1RC2 MH089951). We acknowledge support from VU Amsterdam and the Institute for Health and Care Research.

## Conflict of Interest

None.

## Ethical Standards

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

## Supplementary Material

To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/thg.2016.53>.

## References

- Glasner, T. J., van Beijsterveldt, C. E. M., Willemsen, G., & Boomsma, D. I. (2013). Meerlinggeboorten in Nederland [Multiple births in the Netherlands]. *Nederlands Tijdschrift voor Geneeskunde*, 57, A5962.
- Loos, R. J. F., & Yeo, G. S. H. (2014). The bigger picture of FTO – the first GWAS-identified obesity gene. *Nature Review Endocrinology*, 10, 51–61.
- Mbarek, H., Steinberg, S., Nyholt, D. R., Gordon, S. D., Miller, M. B., McRae, A. F., ... Boomsma, D. I. (2016). Identification of common genetic variants influencing spontaneous dizygotic twinning and female fertility. *The American Journal of Human Genetics*, 98, 898–908.
- Pison, G., Monden, C., & Smits, J. (2015). Twinning rates in developed countries: Trends and explanations. *Population and Developmental Review*, 41, 629–649.
- Wray, N., & Visscher, P. (2007). Population genetics and its relevance to gene mapping. In B. Neale, M. Ferreira, S. Medland & D. Posthuma (Eds.), *Statistical genetics: Gene mapping through linkage and association*. (pp. 87–110). London: Taylor and Francis.

TABLE A1

## Genotype By Twinning Probability

Genotype	DZ yes	DZ no	Genotype probability
FB(G;G)	a	b	<i>a+b</i>
FB(G;A)	c	d	<i>c+d</i>
FB(A;A)	e	f	<i>e+f</i>
	<i>a+c+e</i>	<i>b+d+f</i>	1

## Appendix

In the following Table A1 (based on FSHB), the marginal probabilities, that is, the genotype probabilities (*a+b*, *c+d*, *e+f*) and the spontaneous DZ twinning probability (*a+c+e*) are given in italics. In addition, given the additive model, we have  $(c/(c+d))/(e/(e+f)) = (a/(a+b))/(b/(a+b)) = RR$  (relative risk). The ORs (odd ratios) with FB(A;A) are the reference, are  $(c/d)/(e/f)$  and  $(a/b)/(e/f)$ .

Based on this information, we calculated the probabilities, denoted a to f in Table A2, by simple least squares in R (the R code is available as supplemental material). The haplotype by twinning probability table can be obtained in the same manner. However, it is easier to first construct Table A1 (see Table 1), to calculate  $\text{prob}(\text{FB genotype} \mid \text{DZ outcome})$  and  $\text{prob}(\text{S3 genotype} \mid \text{DZ outcome})$ , and then to use these to approximate  $\text{prob}(\{\text{FB genotype and S3 genotype}\} \mid \text{DZ outcome})$  as  $\text{prob}(\text{FB genotype} \mid \text{DZ outcome}) * \text{prob}(\text{S3 genotype} \mid \text{DZ outcome})$ . This approximation is good, because the effect sizes are relatively

TABLE A2

## Approximate FSHB (FB) and SMAD3 (S3) Haplotypes By Spontaneous Twinning Probabilities

Haplotype	DZ yes	DZ no	Haplotype frequency	Prob(DZ haplotype)
FB(G;G) & S3(C;C)	0.00053	0.04108	<i>0.04162</i>	0.01276
FB(G;G) & S3(C;T)	0.00308	0.26048	<i>0.26357</i>	0.01170
FB(G;G) & S3(T;T)	0.00448	0.41284	<i>0.41732</i>	0.01074
FB(A;G) & S3(C;C)	0.00016	0.01453	<i>0.01469</i>	0.01081
FB(A;G) & S3(C;T)	0.00092	0.09210	<i>0.09302</i>	0.00992
FB(A;G) & S3(T;T)	0.00134	0.14595	<i>0.14729</i>	0.00910
FB(A;A) & S3(C;C)	0.00001	0.00128	<i>0.00130</i>	0.00918
FB(A;A) & S3(C;T)	0.00007	0.00814	<i>0.00821</i>	0.00841
FB(A;A) & S3(T;T)	0.00010	0.01290	<i>0.01300</i>	0.00771
	<i>0.0107</i>	<i>0.9893</i>	1	

Note: The last row contains the marginal spontaneous DZ twinning probability (in italics; *0.0107*). Column 5 contains the conditional probabilities ( $\text{prob}(\text{DZ} \mid \text{haplotype})$ ). As expected, these closely resemble those given in Table 4; column 5).

small. Given these conditional probabilities and  $\text{prob}(\text{DZ outcome} \mid \{\text{FB genotype and S3 genotype}\})$ . Subsequently, given these and the haplotype probabilities (which depend only on the allele frequencies), we calculated the 18 entries of the haplotype x DZ twinning table (see Table A2). The R code we used to obtain the entries in Table A2 is available as supplemental material.