Published in final edited form as:

Trends Genet. 2010 June ; 26(6): 241–243. doi:10.1016/j.tig.2010.03.003.

Detecting genes in complex disease: does phenotype accuracy limit the horizon?

Koon Lee¹ and Stephen Sawcer¹

¹University of Cambridge, Department of Clinical Neuroscience, Addenbrooke's, Hospital, Hills Road, Cambridge, CB2 0QQ, UK

In their recent article, Samuels et al. [1] suggested that diagnostic error rates of just a few percent have such a detrimental effect on power that massive increases in sample size would be required to offset their adverse effects. The authors therefore concluded that the accuracy of phenotyping is a major factor limiting the ability of association studies to identify novel genes in complex traits. We are writing to point out that they have seriously over estimated the influence of this source of error; recalculation shows that even in the worst case scenario diagnostic error rates would have to be higher than 10% before they are likely to have any substantial effect on power.

No diagnostic process is perfect and no matter how careful researchers are it is inevitable that some individuals who are unaffected will be mistakenly labelled as cases and vice versa. The PAWE program [2] allows researchers to calculate the power to detect association in the face of these diagnostic errors. Within the PAWE program error rates are specified in terms of the conditional probabilities $\Pr(\text{observed control} \mid \text{true affected})$ (the proportion of truly affected individuals mistakenly labelled as unaffected) and $\Pr(\text{observed case} \mid \text{true unaffected})$ (the proportion of truly unaffected individuals mistakenly labelled as affected). In using this program it is important to remember that the diagnostic error rate is the proportion of individuals labelled as affected that are actually unaffected i.e. the conditional probability $\Pr(\text{true unaffected} \mid \text{observed case})$ and that in general:

$$\Pr(\text{true unaffected} \mid \text{observed case}) \neq \Pr(\text{observed case} \mid \text{true unaffected})$$

Unfortunately in their analysis of the relationship between power and diagnostic error rate, Samuels et al. [1] seem to have mistakenly used the diagnostic error rate as the input for the PAWE program and have thereby substantially overestimated the impact of diagnostic errors on power. For a given prevalence and diagnostic sensitivity, straight forward algebra enables the correct input for PAWE, $\Pr(\text{observed case} \mid \text{true unaffected})$, to be calculated for any given diagnostic error rate, $\Pr(\text{true unaffected} \mid \text{observed case})$. Here, we reproduce the main figure from Samuels et al. [1] after calculating $\Pr(\text{observed case} \mid \text{true unaffected})$ on the basis of $\Pr(\text{true unaffected} \mid \text{observed case})$ as should have been done (Figure 1). These corrected curves show that diagnostic errors have almost no impact on power when the sample size employed

© 2010 Elsevier Ltd. All rights reserved.

Corresponding Author: Stephen Sawcer, University of Cambridge, Department of Clinical Neuroscience, Addenbrooke's, Hospital, Hills Road, Cambridge, CB2 0QQ, UK, Phone: +44 1223 216073, FAX: +44 1223 336941, sjs1016@mole.bio.cam.ac.uk.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

is inadequate or excessive with respect to the size of effects sought, and only limited impact even when sample size is critical. The figures show that even in the worst case scenario diagnostic error rates as high as 10% reduce power by little more than 20%. By correcting the simple arithmetic error undermining the results presented by Samuels et al. [1] provides a more accurate illustration of how diagnostic error impacts on power which is in line with our previous report on this issue [3]. While it is clear that any diagnostic error reduces power it is also clear that the impact is not as great as has been suggested. In short our answer to the question posed by Samuels et al. [1] in the title of their article would be “yes, but not by as much as you might think”.

Acknowledgments

This work was supported by the Wellcome Trust (084702/Z/08/Z), the Medical Research Council (G0700061), the National Institute of Health (RO1 NS049477) and the Cambridge NIHR Biomedical Research Centre.

References

1. Samuels D, Burn D, Chinnery P. Detecting new neurodegenerative disease genes: does phenotype accuracy limit the horizon? *Trends in Genetics* 2009;Vol.25:486–488. [PubMed: 19819581]
2. Edwards BJ, et al. Power and sample size calculations in the presence of phenotype errors for case/control genetic association studies. *BMC Genet* 2005;6:18. [PubMed: 15819990]
3. Sawcer S. The complex genetics of multiple sclerosis: pitfalls and prospects. *Brain* 2008;131:3118–3131. [PubMed: 18490360]

Prevalence(K)	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.01
PropMissedCases(r)	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001
Cases	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
AlleleFreq	0.500	0.250	0.100	0.500	0.500	0.125	0.125	0.125	0.100	0.100	0.100
GRR	1.3	1.3	1.3	1.2	1.1	1.3	1.3	1.3	1.3	1.2	1.1
p-value	0.0000005	0.0000005	0.0000005	0.0000005	0.0000005	0.0000005	0.0000005	0.0000005	0.0000005	0.0000005	0.0000005

DiagnosticError(g)	power	power	power	power	power	power	power	power	n-error	ratio (N/N ₀)	n-error	ratio (N/N ₀)	n-error	ratio (N/N ₀)
0	1.000000	1.000000	1.000000	1.000000	0.931823	1.000000	1.000000	1.000000	6933	1.000000	14781	1.000000	55902	1.000000
0.1	1.000000	1.000000	1.000000	1.000000	0.792081	1.000000	1.000000	1.000000	8472	1.221982	18117	1.225695	68753	1.229884
0.2	1.000000	1.000000	0.999988	1.000000	0.555183	1.000000	1.000000	1.000000	10610	1.530362	22761	1.539882	86681	1.550589
0.3	1.000000	1.000000	0.998977	0.999916	0.296327	0.999951	0.999940	0.999930	13709	1.977355	29507	1.996279	112776	2.017388
0.4	1.000000	0.999999	0.973115	0.993285	0.113657	0.995682	0.995111	0.994611	18454	2.661763	39858	2.696570	152901	2.735162
0.5	0.999997	0.999226	0.777382	0.881519	0.030216	0.910664	0.904723	0.899819	26275	3.789846	56954	3.853190	219316	3.923223
0.6	0.983860	0.933654	0.340062	0.458395	0.005463	0.521039	0.509473	0.500290	40579	5.853022	88293	5.973412	341308	6.105470
0.7	0.615299	0.433749	0.055109	0.082606	0.000671	0.105810	0.101846	0.098785	71281	10.281408	155710	10.534470	604335	10.810615
0.8	0.060676	0.032686	0.002653	0.003868	0.000057	0.005271	0.005052	0.004886	158423	22.850570	347522	23.511400	1354382	24.227791
0.9	0.000371	0.000223	0.000036	0.000045	0.000004	0.000057	0.000056	0.000054	625743	90.255733	1378540	93.264326	5394922	96.506780
1	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	Infinite	1.00E+15	Infinite	1.00E+15	Infinite	1.00E+15

Figure 1.

Power to detect a genetic association in the context of diagnostic errors. In each example, the probability of affected individuals being classified as controls is 1×10^{-5} . Varying this parameter has negligible impact on power and/or optimal sample size for diseases that are present in less than 10% of the population. (a) Power to detect an association between a common allele (allele frequency = 0.5; GRR = 1.1–1.3 under a multiplicative model) and disease in 20000 cases and 20000 controls with varying degrees of diagnostic error at $P < 5 \times 10^{-7}$. Disease frequency = 0.01. (b) Power to detect an association between alleles of different frequency (0.5, 0.25, 0.1) and disease in 20000 cases and 20000 controls with varying degrees of diagnostic error at $P < 5 \times 10^{-7}$. GRR = 1.3, disease frequency = 0.01. (c) Power to detect an association between an allele (frequency = 0.125, GRR = 1.3) and diseases of different prevalence (0.01, 0.001, 0.0001) in 20000 cases and 20000 controls with varying degrees of diagnostic error at $P < 5 \times 10^{-7}$. (d) Ratio of the number of inaccurately phenotyped cases (n_{error}) to the number of accurately phenotyped cases ($n_{\text{no error}}$) required to detect an association between an allele (frequency = 0.1, varying GRR from 1.1 to 1.3) and a disease (frequency = 0.01) with 95% power at varying degrees of diagnostic error at $P < 5 \times 10^{-7}$. All calculations used PAWE-PH Phenotype Edition [2].