GENETIC TESTING
Volume 11, Number 4, 2007

DOI: 10.1089/gte.2007.0028

On the Testing Load Incurred by Cascade Genetic Carrier Screening for Mendelian Disorders: A Brief Report

MICHAEL KRAWCZAK,[1] AMKE CALIEBE,[1] PETER J.P. CROUCHER,[1,2] and JÖRG SCHMIDTKE[3]

ABSTRACT

One criterion to decide to whom molecular genetic carrier testing should be provided is an individual's carrier risk, taking into account his or her affection status and degree of relatedness to an overt carrier. We have derived formulas to calculate the testing load incurred to a public health system following such a cascade screening strategy. While the testing load turns out to be moderate for individual diseases at meaningful risk thresholds (i.e., 1%–5%), a substantial proportion of the population would have to be tested if all known single gene disorders were to be included in a cascade screening program.

INTRODUCTION

RECENT PROGRESS in medical genetics, in part fuelled by the Human Genome Project, has led to the availability of an increasing number of molecular genetic tests for diagnostic and predictive purposes. A large proportion of the estimated 4000 to 6000 single-gene disorders are already amenable to DNA sequence analysis (www.genetests.org, www.orpha.net). Although most diseases in this category are individually rare (prevalence $\leq 5 \times 10^{-4}$), their estimated cumulative frequency of up to 5% renders carrier detection for Mendelian disorders a major challenge to the health care system (van Weely and Leufkens, 2004). The potential testing load incurred by the comprehensive molecular genetic screening of many, if not all, monogenic disorders has raised concerns about the costs and benefits, particularly in publicly financed health systems (Burke and Zimmern, 2005). Consequently, there is a strong demand for a proper evaluation of the validity and utility of carrier tests, particularly to guide test providers and regulators in their decision making process as to which tests to implement in practice (Sanderson *et al*., 2005; www.cdc.gov/genomics/gtesting). One plausible criterion for a rational delineation of testing efforts would be the probability with which an individual carries a disease-associated genotype, i.e., the individual's genetic risk, based upon its degree of relatedness to patients affected by the disease in question. We have previously assessed the efficiency and efficacy of such a cascade screening approach (Krawczak *et al*., 2001). Surprisingly, however, the disease-specific and overall genetic testing load incurred to the public health system by the adoption of a cascade screening policy have apparently not yet been formally assessed in the scientific literature. We have therefore derived appropriate formulas for performing the corresponding calculations and applied them to prevalence and penetrance figures representative of archetypical single-gene disorders.

MATERIALS AND METHODS

In order to assess the genetic testing load for a public health system, it is sufficient to consider what proportion $\pi(r_0)$ of the unaffected population is closely related to an overt (i.e., affected) mutation carrier, and therefore has a carrier risk above a given threshold r_0. For the sake of simplicity, we will confine our considerations to nonoverlapping generations, assuming that at-risk individuals are identified only through their relatedness to an index patient in the same generation. This simplification is unlikely to impact much upon the validity of our results for dominant conditions. For recessive mutations, in contrast, the proportion of individuals facing a certain carrier risk may be underestimated if only same-generation patients are alluded to. However, since the overall testing load is also likely to be reduced when information from preceding or succeeding

[1]Institut für Medizinische Informatik und Statistik, Christian-Albrechts-Universität, Kiel, Germany.
[2]Department of Environmental Science, Policy & Management, College of Natural Resources, University of California, Berkeley, California.
[3]Institut für Humangenetik, Medizinische Hochschule, Hannover, Germany.

generations is taken into account, the net bias for recessive conditions is likely to be small as well.

If the population of interest is large enough, panmictic and of constant size, then the number of relatives of an index patient can be assumed to follow a Poisson distribution, with expected values of unity for siblings and of 2^n for unilateral n^{th} degree relatives ($n > 1$) in the same generation as the patient (e.g., $2^2 = 4$ first-degree cousins, $2^3 = 8$ second degree cousins, etc.). For a single carrier, the total number of relatives of at least n^{th} degree therefore follows a Poisson distribution with expected value equal to $x_n = 1 + 4 + ... + 2^n = 2^{n+1} - 3$. Now, let p be the prevalence of overt mutation carriers in the population. In order to determine the proportion π_n of the population who are at least n^{th} degree relatives of at least one overt carrier, we use the following mathematical model. Let N be the total population size and let $k = p \cdot N$ denote the number of overt carriers. We now assume that all relatives of the k overt carriers are successively drawn, with replacement, from the N-k individuals who are not overt carriers themselves. Drawing with replacement allows for the fact that moderately distant relatives of different carriers may coincide. The error introduced by the fact that an individual can be drawn twice as a relative of one and the same carrier is negligible because x_n is small compared to N for most practically relevant values of n. Under these assumptions, π_n can be approximated by

$$\pi_n \approx 1 - e^{-xn \cdot \frac{p}{1-p}} \quad (1)$$

For a detailed derivation of formula 1, see Appendix 1.

For a dominant mutation, the prior carrier risk r_n of an n^{th} degree, same-generation relative of an overt heterozygous carrier equals $4^{-n+1/2}$ (i.e., 1/2 for a sibling, 1/8 for cousins, etc.). If ψ denotes the penetrance of the mutation in heterozygous carriers, then the expected proportion of unaffected n^{th} degree relatives is

$$\pi_n^* = (\pi_n - \pi_{n-1}) \cdot (1 - \psi \cdot r_n) \quad (2)$$

For a given risk threshold r_0, this implies

$$\pi(r_0) = \sum_{i=1}^{n_0} \pi_i^* \quad (3)$$

where n_0 is the largest integer n for which r_n corresponds to a posterior risk of at least r_0, i.e.

$$n_0 = \max\left(n{:}r_n \geq \frac{r_0}{1 - \psi(1 - r_0)}\right) \quad (4)$$

In the case of recessive mutations, $r_1 = 3/4$ and $r_n = 4^{-n+1}$ for $n \geq 2$. Then, for a given risk threshold r_0,

$$\pi(r_0) = \pi_{n_0} \quad (5)$$

where n_0 is the largest integer n for which r_n is at least as large as r_0.

The utility of a genetic screening program can be assessed by its efficacy R, i.e., the proportion of carriers detected by the program, and efficiency E, i.e., the number of individuals that have to be tested per detected carrier (Krawczak *et al.*, 2001).

TABLE 1. TESTING LOAD FOR A DOMINANT MUTATION WITH REDUCED PENETRANCE

Ψ	p	r_0	$\pi(r_0)$	$R(r_0)$	$E(r_0)$
0.8	5×10^{-4}	0.01	2.098×10^{-3}	0.800	21.0
		0.05	3.000×10^{-4}	0.400	6.0
	5×10^{-5}	0.01	2.100×10^{-4}	0.800	21.0
		0.05	3.000×10^{-5}	0.400	6.0
0.5	5×10^{-4}	0.01	6.170×10^{-3}	0.625	19.8
		0.05	2.248×10^{-3}	0.500	9.0
	5×10^{-5}	0.01	6.186×10^{-4}	0.625	19.8
		0.05	2.250×10^{-4}	0.500	9.0
0.1	5×10^{-4}	0.01	6.420×10^{-3}	0.125	11.4
		0.05	2.448×10^{-3}	0.100	5.4
	5×10^{-5}	0.01	6.438×10^{-4}	0.125	11.4
		0.05	2.450×10^{-4}	0.100	5.4

Ψ, penetrance, *p*, prevalence of overt heterozygous carriers, r_0, risk threshold, $\pi(r_0)$, proportion of the population that needs to be screened; $R(r_0)$: screening efficacy, i.e., the proportion of same-generation carriers detected; $E(r_0)$: screening efficiency, i.e., number of individuals tested per same generation carrier detected.

Within the present mathematical framework, $R(r_0)$ can be calculated as

$$R(r_0) = h^{-1} \cdot \sum_{i=1}^{n_0} (\pi_i - \pi_{i-1}) \cdot r_i^* \quad (6)$$

where h is the population frequency of unaffected carriers and r_i^* equals the posterior carrier risk of unaffected i^{th} degree relatives of a patient, i.e.,

$$r_i^* = \frac{r_n \cdot (1 - \psi)}{(1 - \psi \cdot r_n)} \quad (7)$$

for dominant diseases, and $r_i^* = r_i$ for recessive diseases. Efficiency $E(r_0)$ equals the ratio of $\pi(r_0)$ and $h \cdot R(r_0)$.

RESULTS AND DISCUSSION

Tables 1 and 2 summarize the application of the above formulas to archetypical prevalence and penetrance figures per-

TABLE 2. TESTING LOAD FOR A RECESSIVE MUTATION WITH COMPLETE PENETRANCE

p	r_0	$\pi(r_0)$	$R(r_0)$	$E(r_0)$
5×10^{-4}	0.01	1.440×10^{-2}	7.672×10^{-3}	42.0
	0.05	6.482×10^{-3}	6.980×10^{-3}	20.8
5×10^{-5}	0.01	1.449×10^{-3}	2.430×10^{-3}	42.2
	0.05	6.498×10^{-4}	2.210×10^{-3}	20.8
5×10^{-6}	0.01	1.500×10^{-4}	7.686×10^{-4}	42.2
	0.05	6.500×10^{-5}	6.988×10^{-4}	20.8

p, prevalence of heterozygous carriers, r_0, risk threshold, $\pi(r_0)$, proportion of the population that needs to be screened; $R(r_0)$: screening efficacy, i.e., the proportion of same-generation carriers detected; $E(r_0)$: screening efficiency, i.e., number of individuals tested per same generation carrier detected.

taining to human genetic disorders. For conditions with a prevalence of up to 5×10^{-4}, only a small fraction of the population would have to be tested for carrier assessment at practically meaningful risk thresholds. If the cumulative prevalence of these conditions is taken into account, however, it turns out that a substantial proportion of the population would eventually have to be tested, particularly if low-penetrance dominant mutations were included. This conclusion was not unexpected, bearing in mind earlier considerations of the human genetic load (Vogel and Motulsky, 1997).

The above notwithstanding, it is important to realize that the acceptability of a genetic screening program is unlikely to dependent upon the load incurred to society alone. We have emphasized before (Krawczak *et al*., 2001) that the cost effectiveness of a screening program is a function of both, its efficiency and the material and immaterial costs associated with carrying out a single diagnosis (c_d) or managing an undetected case (c_m). A screening program is cost-effective if $E < c_m/c_d$. Inspection of Tables 1 and 2 reveals that cascade genetic screening may be cost effective for dominant conditions, particularly if these are of low penetrance and if the screening efforts are targeted at high-risk individuals. For recessive conditions, however, cost effectiveness is unlikely and the efficacy of cascade screening appears too low for it to be justified in practice, even if the overall testing load may seem manageable.

Given current technologies and health care frameworks, comprehensive genetic testing at a population-wide level appears impracticable. On the other hand, predictive testing in the form of cascade screening is recognized as requiring restriction to particular settings. We hope that the computational tools provided here and in a previous publication (Krawczak *et al*., 2001) will prove useful to decide if and when such undertakings are politically and economically sensible.

ACKNOWLEDGEMENTS

This study was partly supported by the German Ministry of Science and Education ('BMBF') through an NGFN SMP-GEM grant to M.K. (01GS0426).

J.S. was supported by EUROGENTEST (www.eurogentest.org).

REFERENCES

Burke W, Zimmern RL (2005) Ensuring the appropriate use of genetic tests. Nat Rev Genet 5:955–959.

Krawczak M, Cooper DN, Schmidtke J (2001) Estimating the efficacy and efficiency of cascade genetic screening. Am J Hum Genet 69:361–370.

Sanderson S, Zimmern R, Kroese M, Higgins J, Patch C, Emery J (2005) How can the evaluation of genetic tests be enhanced? Lessons learned from the ACCE framework and evaluating genetic tests in the United Kingdom. Genet Med 7:495–500.

Vogel F, Motulsky A (1997) Human genetics, 3rd ed. Problems and approaches. Springer, Berlin, pp. 559–568

van Weely S, Leufkens HGM (2004) Orphan diseases. In: Kaplan W, Laing R (eds) Priority medicines for Europe and the world–A public health approach to innovation. World Health Organization, Geneva, pp. 95–100.

Address reprint requests to:
Michael Krawczak
Institut für Medizinische Informatik und Statistik
Christian-Albrechts-Universität
Brunswiker Strasse 10
24105 Kiel
Germany

E-mail: krawczak@medinfo.uni-kiel.de

Appendix 1.

Under the mathematical model used:

$$\text{(A1)}\quad \pi_n \approx E\left[1 - \left(1 - \frac{1}{N-k}\right)^{T_n}\right] = 1 - E\left[\left(1 - \frac{1}{N-k}\right)^{T_n}\right]$$

where $E[\cdot]$ denotes expectation and T_n follows a Poisson distribution with expected value $k \cdot x_n$. Substituting z for the base of T_n in the above formula yields

$$\text{(A2)}\quad E[z^{Tn}] = e^{-k\cdot x_n}\sum_{j=0}^{\infty} z^j \cdot \frac{(k\cdot x_n)^j}{j!} = e^{-k\cdot x_n}\sum_{j=0}^{\infty}\frac{(z\cdot k\cdot x_n)^j}{j!} = e^{-k\cdot x_n}\cdot e^{z\cdot k\cdot x_n} = e^{-k\cdot x_n\cdot(1-z)}$$

which completes the proof of formula 1 in the main text, recalling that

$$\text{(A3)}\quad -k\cdot x_n\cdot(1-z) = -x_n\cdot\frac{k}{N-k} = -x_n\cdot\frac{p}{1-p}.$$