The Influence of the Wahlund Effect on the Consanguinity Hypothesis: Consequences for Recessive Disease Incidence in a Socially Structured Pakistani Population

Andrew D.J. Overall

Division of Biology, University of Brighton, Brighton, UK

Key Words

Consanguinity · Population subdivision · Wahlund effect · Recessive disease

Abstract

Background/Aims: Standard population genetic theory predicts that the relative risk of inheriting recessive disorders between consanguineous and non-consanguineous populations can be manyfold. However, it is rarely considered that consanguineous populations might be composites of socially defined endogamous and genetically differentiated subpopulations. A recent study of a British Pakistani population found evidence to suggest that extended families (*biraderi*) could contribute significantly to excessive homozygosity over that contributed by consanguinity. This study sets out to illustrate the potential of cryptic population substructure (the Wahlund effect) to contribute to recessive disease incidence in populations with complex social structure. **Method:** Population parameter estimates were drawn from a recent study of the British Pakistani population along with allele frequency estimates of nine recessive inborn errors of metabolism. The relative contribution of consanguinity and *biraderi* endogamy to recessive disease incidence was predicted. **Results:** Population substructure of the magnitude estimated from studies of *biraderi* endogamy are sufficient to significantly contribute to the incidence of recessive disorders within consanguineous populations. **Conclusions:** Because *non-consanguineous* couples have a higher risk of

sharing the same recessive mutation in a substructured population relative to a non-substructured population, the health benefits of avoiding consanguinity in these situations is likely to be less pronounced than the standard consanguinity hypothesis predicts.

Copyright © 2008 S. Karger AG, Basel

Introduction

Standard population genetics theory provides a relationship between the degree of relatedness within the parental generation and the probability of observing particular genotypes in the offspring generation [1, p 263]. This is sometimes referred to as the consanguinity hypothesis because as the inbreeding coefficient, f_b , increases, the probability of observing individuals homozygous at any particular locus increases, which has clear implications for the population frequency of recessive disorders. On the basis of this simple relationship, the consanguinity hypothesis is often used to explain the high incidence of recessive disorders in populations known to practice consanguinity. For example, several UK studies have identified a high incidence of recessive disorders within populations originating from Pakistan [2–5]. One of the consequences of these UK-based studies has been recent public calls in the British media, to discourage consanguineous marriages amongst British Pakistani individuals, which in turn has prompted criticism from both

community leaders and researchers arguing that such a relationship is overly simplistic [6].

The British Pakistani communities have levels of consanguinity ranging between 50 and 70% first cousin marriages [5, 7]. One particular study of the Pakistani population living in Birmingham, UK, found that autosomal recessive diseases were 16-fold higher in offspring born to consanguineous parents compared to non-consanguineous couples [8]. On this basis it might seem incontrovertible that avoiding consanguinity would bring clear and significant health benefits to these communities. However, there are also incidences of high autosomal recessive rates amongst non-consanguineous Asian populations. For example, Young & Clark [3] compared the incidence of lethal malformations between two groups broadly categorized as 'White' and 'Asian' living in Leicester, UK, where the majority of the 'Asian' individuals were of Indian descent. This study found a significantly higher rate of recessive disorders amongst the Asian group with the highest rate amongst the offspring of non-consanguineous Hindu parents mostly descending from Gujarat. Terry et al. [2, 9] also identified the Indian population of Birmingham as having the highest stillbirth and perinatal mortality amongst several ethnic groups studied, despite the absence of consanguinity. Although a genetic predisposition was considered a likely explanation for this observation, no suggestion was put forward to explain the apparent excess of mortality. One possibility not explored is the presence of caste endogamy, of which indirect evidence has been obtained from a genetic study of the Indian community of Nottingham, UK [10]. When a population is substructured into numerous cryptic subpopulations, such as caste, excess homozygosity greater than Hardy-Weinberg expectations can occur. This is the Wahlund effect [11], which can also potentially increase the incidence of recessive disorders.

Population subdivision has been identified in the British Asian population amongst families descending from both India and Pakistan. Within Pakistani Muslim communities, the majority of marriages occur within broad, extra-familial network groups, sometimes referred to as *biraderi*. Unfortunately, *biraderi* are difficult to define, for example Wakil [12] states that: 'The boundaries of *biraderi* are undefined and usually very hazy. The size of the *biraderi* is as large as the distance at which one can recognise one's relatives'. Despite the cryptic nature of *biraderi*, they nevertheless provide a plausible mechanism for generating population stratification beyond that of the more readily identifiable caste/*zat* substructuring; as was observed in a genetic survey of the Pakistani population of Nottingham, UK [13].

There are possible consequences to the genetic health of substructured populations in that, although the average frequencies of deleterious alleles might appear to be at a low frequency in the total British Pakistani population, there are likely to be some *biraderi* that have a higher than average frequency of a specific recessive allele, and other *biraderi* with a lower than average frequency. However, what is not clear from the standard explanation of risk incorporating the consanguinity hypothesis is that, if populations like the British Pakistani are substructured through having endogamous kin-networks, just how much of the recessive disease incidence is contributed to by consanguinity and how much of a difference in recessive incidence could we expect if consanguinity was avoided?

It is possible that population substructure of the magnitude observed in the British Pakistani population is sufficient to contribute significantly to the incidence of recessive disease observed and could, for rare alleles (e.g., $q < 0.01$), considerably mitigate the relative risk of recessive disease between consanguineous and non-consanguineous families. This notion is at odds with the prevailing consanguinity hypothesis which implies marked reductions in incidence in the absence of consanguinity. Avoiding consanguinity in a substructured population would not, therefore, necessarily result in significant reductions in incidence. The relevance of this model is that it suggests the benefits of a more family-orientated approach to addressing high incidence of recessive disorders within populations displaying complex social structuring, for the simple reason that large extended families could in certain circumstances harbour mutation frequencies showing considerable variability from each other. Furthermore, evidence for population substructure would add additional drive to balance the over-simplified and socially disruptive theories on the genetic consequences of close-kin unions.

Methods

Data obtained from Hutchesson et al. [5] and Overall et al. [13] are used to illustrate the potential relative contributions of consanguinity and substructure to recessive disease incidence. Hutchesson et al. [5] published disease frequencies of ten recessive inborn errors of metabolism (IEM) observed within populations of NW European and Pakistani origin living in the West Midlands, UK. One of the ten disorders, MCADD, was absent from the Pakistani population and so is not considered further in this analysis. This study also reported that the Pakistani population was composed of around 70% consanguines, where the average inbreeding coefficient was $f_I = 0.0686$, slightly higher than that of

Fig. 1. a The expected frequency of inborn errors of metabolism in 1st cousin offspring ($= q(1/16 + 15q/16)$, solid line) and non-consanguineous offspring ($= q^2$, bold dashed line). The ratio (R) of these two expected frequencies is given by the feint dashed line. **b** The expected frequency of inborn errors of metabolism in a substructured population in 1st cousin offspring ($= q(1/16 + 15/16(F_{ST} + (1 - F_{ST})q))$, $F_{ST} = 0.03$, solid line) and non-consanguineous offspring ($= q(F_{ST} + (1 - F_{ST})q)$, $F_{ST} = 0.03$, bold dashed line). The ratio (R) of disease incidence between offspring of 1st

cousins and offspring of non-consanguines given by feint dashed line. The predicted allele frequencies (q) given in Hutchesson et al. [5], are indicated by the arrows, where the disorders, in order of increasing expected disease frequency, are: A: hypoeroxaluria type 1 and phenylketonuria; B: galactosaemia; C: mucopolysaccharidosis, San Fillipo disease type C, Niemann Pick disease type C, and non-ketonic hyperglycaemia; D: cystinosis; E: tyrosinaemia type 1.

first cousin offspring. This value was calculated using family history data collected as part of the five-year study of 956 babies of British Pakistani origin, born in Birmingham, UK [8, 14]. Using these data, the authors estimated the underlying allele frequencies for each of the disorders for the NW European and Pakistani populations. The small population proportions of the individual diseases means that that there is considerable error around these allele frequency estimates, hence the mean values are used for illustrative purposes only. However, assuming these frequencies, the expected relative risk of homozygosity between consanguineous offspring (1st cousin offspring, $f_I = 0.0625$) and non-consanguineous offspring within the Pakistani population can be estimated where each of the underlying disease alleles is denoted a , with a frequency of q , [1, p 263]:

$$\frac{\Pr(aa|\text{consanguineous})}{\Pr(aa|\text{non-consanguineous})} = \frac{q(f_I + (1 - f_I)q)}{q^2}. \quad (1)$$

Extending this logic to consider the expected relative risk of homozygosity between consanguines and non-consanguines within a substructured population, this ratio becomes (the ratio of eq. 2 and eq. 1(a) in [10]):

$$\frac{\Pr(aa|\text{consanguineous, substructured})}{\Pr(aa|\text{non-consanguineous, substructured})} = \frac{q(f_I + (1 - f_I)(F_{ST} + (1 - F_{ST})q))}{q(F_{ST} + (1 - F_{ST})q)}. \quad (2)$$

A comparison of these two ratios demonstrates the importance of population substructure to the incidence of recessive disorders relative to consanguinity.

Results

Figure 1a presents the expected probability of homozygosity in consanguineous and non-consanguineous populations and the ratio of these expectations (ratio (1)) with increasing allele frequency. The arrows indicate the frequencies of the nine IEM, illustrating that the expected discrepancy between the incidence of the disorders ranges from hyperoxaluria type 1 and phenylketonuria (A in fig. 1a), which are approximately 46 times more prevalent amongst the consanguines than non-consanguines, to tyrosinaemia type 1 (E in fig. 1a), which is approximately 10 times more prevalent; reflecting that this discrepancy is greater when the underlying allele frequency is rarer.

Figure 1b shows the expectation for a population substructured at a magnitude towards the upper region of the plausible range ($F_{ST} = 0.03$ [13]). It is clear that, unlike the unstructured population ($F_{ST} = 0$), where the ratio is

Fig. 2. **a** Plot showing the relationship between the expected ratio (R) of homozygosity amongst consanguines relative to non-consanguines in a population substructured at varying degrees ($F_{ST} = 0.001$ – 0.06) with changing allele frequency. **b** Plot showing the relationship between the expected ratio (R) of homozygosity amongst the offspring generation of consanguines relative to non-consanguines in a substructured population with varying degrees of consanguinity ($f_I = 0.001$ – 0.06) with changing allele frequency.

highly variable (fig. 1a, faint dashed line), the expected ratio of disease incidence for the offspring of 1st cousins relative to offspring of non-consanguines in a substructured population is approximately equal to 3-fold across all recessive disorders (fig. 1b, faint dashed line).

Figure 2a shows how this relationship changes with varying magnitudes of substructure. It appears that even with quite moderate levels of substructure: $F_{ST} \geq 0.01$, the probability of homozygosity when the recessive allele is relatively common, $q > 0.01$, is only slightly higher for consanguines relative to non-consanguines (using ratio (2)). When the alleles are very rare, the effect of substructure in reducing this discrepancy appears to be substantial. For example, individuals homozygous for an allele at frequency $q = 0.0001$ are expected to be around 60 times more prevalent amongst the offspring of consanguines than non-consanguines when substructure is very low, $F_{ST} = 0.001$, but reduce to around seven times as prevalent when substructure increases to $F_{ST} = 0.01$. By way of comparison, figure 2b illustrates the relatively modest influence on this ratio of increasing the magnitude of consanguinity within a substructured population (as the differing scales on the y-axes demonstrate in fig. 2a, b). The relationship between the parameters f_I and F_{ST} on the ratio of expected homozygosity is illustrated in figure 3 and shows that, in the absence of substructure, f_I has a linear relationship with this ratio (R). However, this ratio falls

Fig. 3. Plot showing the relationship between the expected ratio (R) of homozygosity amongst the offspring generation of consanguines relative to non-consanguines in a substructured population with varying magnitudes of f_I and F_{ST} .

in an exponential fashion with increasing substructure, which illustrates the disproportional influence on this ratio of F_{ST} over f_I .

Conclusion

There has been recent public debate on the health consequences of consanguinity; in particular concern has been raised by British government ministers reflecting on the high incidence of birth defects amongst the British Pakistani communities [15, 16]. It is also clear, however, that despite the detriment to genetic health, consanguinity is an important and in many ways beneficial social tradition that should not be discouraged simply on the grounds of increased, but still low, absolute risk of recessive disorders. It has, for example, been suggested that the benefit the family gains through consanguineous marriages might well outweigh the genetic costs [3].

The detrimental potential of consanguinity in more realistic scenarios, where population substructure also influences the relationship between deleterious recessive mutation frequencies and the frequencies of individuals homozygous for these mutations, was explored in further detail. From the analysis presented in figure 1, it appears that within substructured populations, the additional probability of being homozygous due to having consanguineous parents is not as detrimental as the additional probability owing to the population being substructured.

This result is consistent with the finding in the context of forensic DNA analysis, that the probability of observing a matching homozygous genotype (the match probability) is disproportionately influenced by population substructure, with consanguinity contributing much less to the match probability [17]. Consequently, when a population is substructured, the health benefits expected from avoiding consanguinity might not justify the social/cultural costs.

When the magnitudes of f_I and F_{ST} are equivalent, population substructure is likely to have a greater influence on causing deviations from Hardy-Weinberg expectations than consanguinity. This expectation is relevant to future debates on the health consequences of consanguinity in populations such as the British Pakistani population, which have complex social structures built around traditional marriage practices and endogamous extended family networks. In addition to suggesting a contributory role for substructure in the consanguinity hypothesis, this study has also found the current literature to be lacking in crucial information necessary to rigorously test this hypothesis; evidence for population substructure within consanguineous populations, such as the British Pakistani population, is presently limited. Future studies quantifying the roles of *biraderi* and caste in such populations would be of great benefit in addressing the actual detrimental contribution of consanguinity to genetic health.

References

- Hartl DL, Clark AG: Principles of Population Genetics, ed 4. Sunderland, Sinauer, 2007.
- Terry PB, Condie RG, Setttee RS: Analysis of ethnic differences in perinatal statistics. *BMJ* 1980;281:1307–1308.
- Young ID, Clarke M: Lethal malformations and perinatal mortality – A 10 year review. *BMJ* 1987;295:89–91.
- Chitty LS, Winter RM: Perinatal mortality in different ethnic groups. *Arch Dis Child* 1989;64:1036–1041.
- Hutchesson ACJ, Bunday S, Preece MA, Hall SK, Green A: A comparison of disease and gene frequencies of inborn errors of metabolism among different ethnic groups in the West Midlands, UK. *J Med Genet* 1998;35: 366–370.
- Ahmad WIU: Reflections on the consanguinity and birth outcome debate. *J Public Health Med* 1994;16:423–428.
- Darr A, Modell B: The frequency of consanguineous marriage among British Pakistanis. *J Med Genet* 1988;25:186–190.
- Bunday S, Alam H: A five-year prospective study of the health of children in different ethnic groups, with particular reference to the effect of inbreeding. *Eur J Hum Genet* 1993;1:206–219.
- Terry PB, Bissenden JG, Condi RG, Mathew PM: Ethnic differences in congenital malformations. *Arch Dis Child* 1985;62:866–879.
- Overall ADJ, Nichols RA: A method for distinguishing consanguinity and population substructure using multilocus genotype data. *Mol Biol Evol* 2001;18:2048–2056.
- Wahlund S: Composition of populations and correlation appearances viewed in relation to the studies of inheritance. *Hereditas* 1928; 11:65–106.
- Wakil PA: Explorations into kin-networks of the Punjabi society: A preliminary statement. *J Marriage Family* 1970;32:700–707.
- Overall ADJ, Ahmad M, Thomas MG, Nichols RA: An analysis of consanguinity and social structure within the UK Asian population using microsatellite data. *Ann Hum Genet* 2003;67:525–537.
- Modell B, Darr A: Genetic counselling and customary consanguineous marriage. *Nat Rev Genet* 2002;3:225–229.
- Dyer O: MP is criticised for saying that marriage of first cousins is a health problem. *BMJ* 2005;331:1292.
- Gadher D, Morgan C, Oliver J: (2008). Minister warns of 'inbred' Muslims. *The Sunday Times* <http://www.timesonline.co.uk/tol/news/politics/article3342040.ece>, accessed February 2008.
- Ayres KL, Overall ADJ: Allowing for within-subpopulation inbreeding in forensic match probabilities. *Forensic Sci Int* 1999;103:207–216.