Modelling population genetic screening in rare neurodegenerative diseases

Thomas P Spargo MSc1,*, Alfredo Iacoangeli PhD1,2,3,*,#, Mina Ryten PhD4,5,6, Francesca Forzano MD7, Neil Pearce PhD8, and Ammar Al-Chalabi PhD1,9, #

1Maurice Wohl Clinical Neuroscience Institute, King’s College London, Department of Basic and Clinical Neuroscience, London, UK
2Department of Biostatistics and Health Informatics, King’s College London, London, UK
3NIHR Maudsley Biomedical Research Centre (BRC) at South London and Maudsley NHS Foundation Trust and King’s College London, London, UK
4Genetics and Genomic Medicine, Great Ormond Street Institute of Child Health, University College London, London, UK
5NIHR Great Ormond Street Hospital Biomedical Research Centre, University College London, London, UK
6Department of Clinical Genetics, Great Ormond Street Hospital, London, UK
7Department of Clinical Genetics, Guy’s and St Thomas NHS Foundation Trust, London, UK
8Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK.
9King’s College Hospital, Bessemer Road, London, SE5 9RS, UK.

*co-first author
#correspondence should be addressed to alfredo.iacoangeli@kcl.ac.uk and ammar.al-chalabi@kcl.ac.uk

Abstract

Importance: Genomic sequencing enables rapid identification of a breadth of genetic variants. For clinical purposes, sequencing for small genetic variations is considered a solved problem, while challenges remain for structural variants given the lower sensitivity and specificity. Interest has recently risen among governing bodies in developing protocols for population-wide genetic screening. However, usefulness is constrained when the probability of being affected by a rare disease remains low despite a positive genetic test. This is a common scenario in neurodegenerative disorders. The problem is recognised among statisticians and statistical geneticists but less well understood by clinicians and researchers who will act on these results, and by the general public who might access screening services directly without the appropriate support for interpretation.

Observations: We explore the probability of subsequent disease following genetic screening of several variants, both single nucleotide variants (SNVs) and larger repeat expansions, for two neurological conditions, Huntington’s disease (HD) and amyotrophic lateral sclerosis (ALS), comparing with screening for phenylketonuria which is well established. The risk following a positive screening test was 0.5% for C9orf72 in ALS and 0.4% for HTT in HD, when testing repeat expansions for which the test had sub-optimal performance (sensitivity=99% and specificity=90%), and 12.7% for phenylketonuria and 10.9% for ALS SOD1, when testing pathogenic SNVs (sensitivity=99.96% and
specificity=99.95%). Subsequent screening confirmation via PCR for C9orf72 led to a 2% risk of developing ALS as a result of the reduced penetrance (44%).

Conclusions and Relevance: We show that risk following a positive screening test result can be strikingly low for rare neurological diseases. Accordingly, to maximise the utility of screening, it is vital to prioritise protocols of very high sensitivity and specificity, careful selection of markers for screening, giving regard to clinical interpretability, actionability, high penetrance, and secondary testing to confirm positive findings.

Keywords: genetic screening, genomic, next generation sequencing, health informatics, mathematical model, Bayesian probability tools
## Introduction

Single nucleotide variants (SNVs) and small insertions and deletions (indels) are now genotyped in next-generation sequencing data with over 99% accuracy (1-3), while genotyping larger, structural, variants is often less reliable (4). For clinical purposes, the validity of sequencing for small genetic variations is considered a solved problem, while challenges remain for structural variants.

Widened availability of sequencing data has favoured our understanding of the role of genetic factors in various phenotypes (5-8). Genetic testing has therefore become valuable for healthcare and gained popularity among consumers who can it directly, without advice from trained clinicians to guide interpretation of results (9-11).

Interest has recently risen among governing bodies in developing protocols for population-wide genetic screening; such initiatives are being rolled out in the UK and considered in the US (12-18). Genetic screening involves testing a population for genetic variants indicative of risk for specific diseases to identify people with either higher predisposition of developing that disease or the potential to pass it on to their offspring. This approach utilises modern sequencing techniques to evaluate multiple genes associated with selected traits. In contrast ‘targeted’ tests are those performed because of some suggestion that a person may harbour disease variants (e.g., symptoms or family history of disease). Although screening is relevant to liabilities ranging between monogenic and polygenic (cf. (16, 19)), we focus here on screening for pathogenic variants with monogenic associations with rare diseases, particularly as applied to neurodegenerative disorders.

Although no widespread implementation of genetic screening protocols currently exist internationally, comparable metabolic screening, testing neonates for metabolite markers of metabolic diseases, is routine in many countries (17, 20, 21). Positive metabolic tests are typically validated with secondary testing, including targeted genetic tests (9, 22, 23).

The utility of genetic screening can be assessed by the extent to which action can be taken following a positive test: its actionability (24, 25). One key tenet of actionability is the probability of having or later developing a disease following a positive test. Yet post-test disease probability can be strikingly low where disease risk prior to testing is low, as would be for population screening for rare diseases (26).

Bayesian inference, which is routine within clinical decision making (27), can be used to understand post-test (‘posterior’) disease risk. Under this logic, disease probability following a test can be inferred given existing knowledge of the probability of other relevant events. Key considerations to understanding post-test disease risk, beyond pre-test (also known as ‘prior’) disease risk, include the genetic marker penetrance, its frequency among people displaying disease symptoms, and the sensitivity and specificity of the test (analytic validity). This reasoning is therefore highly relevant to screening for rare neurodegenerative diseases, for which genetic causes are typically rare variants of variable penetrance (28).

This article overviews important considerations for genetic screening of rare disorders and it presents several case studies focused on neurodegenerative diseases. Considering conditional probability in medical decision making is not novel but these concepts must be
emphasis in the genomic medicine era, since many results that could be obtained within
large-scale indiscriminate testing of genetic variation across a population will not be
actionable and may be misinterpreted. We modelled genetic screening for Huntington’s
disease (29), HD, and amyotrophic lateral sclerosis (30), ALS, using Bayesian logic to examine
the probability of disease following a positive test result for a dichotomous ALS and HD
genetic markers. We additionally modelled screening for phenylketonuria (PKU) (31), to
compare genetic and metabolic screening.

Methods:
Bayesian framework
We use Bayesian logic to calculate the probability of having or subsequently manifesting
disease D following a test result indicating presence or absence of marker M (the genetic
variant). M is associated with increased liability of D, and positive test result T indicates
presence of M, while negative test result T’ indicates its absence, denoted M’. Disease risk
following a positive result is denoted \( P(D|T) \), using \( P(D|T’) \) for a negative result.
Supplementary Materials 1 summarises the underlying logic. We assume that all model
parameters represent binary events. This was a necessary simplification of reality as, for
example, disease severity is not considered.

We estimate \( P(D|T) \) and \( P(D|T’) \) using the following input parameters:
- \( P(D) \), probability of a person having or later manifesting disease D prior to testing
- \( P(M|D) \), frequency of marker M among those affected by D
- \( P(D|M) \), penetrance, probability of having or later manifesting D for people
  harbouring M
- \( P(T|M) \), sensitivity (true positive rate) of the testing procedure for detecting M
- \( P(T'|M') \), specificity (true negative rate) of the testing procedure for identifying the
  absence of M

Bayes theorem is applied to derive the total probability of harbouring disease marker M,

\[
P(M) = \frac{P(D) \times P(M|D)}{P(D|M)},
\]

and of disease D manifesting given the absence of M,

\[
P(D|M') = \frac{P(D) \times (1 - P(M|D))}{(1 - P(M))}.
\]

We next calculate the total probability of positive test result T, \( P(T) \), according to the
sensitivity and specificity of the test and the probabilities of M being present versus absent:

\[
P(T) = P(T|M) \times P(M) + (1 - P(T'|M')) \times (1 - P(M)).
\]

Bayes theorem is then used to derive the probabilities of M being present after positive,
\[ P(M|T) = \frac{P(M) \times P(T|M)}{P(T)}, \]

and negative, 

\[ P(M|T') = \frac{P(M) \times (1 - P(T|M))}{(1 - P(T))}, \]

results.

The probabilities of manifesting disease \(D\) (which has conditional independence from \(T\) when considering \(M\)) after receiving positive test result \(T\), 

\[ P(D|T) = P(D|M) \times P(M|T) + P(D|M') \times (1 - P(M|T)), \]

and negative test result \(T'\),

\[ P(D|T') = P(D|M) \times P(M|T') + P(D|M') \times (1 - P(M|T')), \]

can then be determined.

Case studies

The Project MinE (32), ALS variant server (33), ClinVar (34), and gnomAD databases were searched alongside the previous databases and next-generation sequencing tool benchmarking reports to obtain data for the included case studies. Most input parameters were defined using data from published literature and online databases. Sensitivity \(P(T|M)\) and specificity \(P(T'|M)\) were defined by performance benchmarks for variant calling with state-of-the-art genomic sequencing techniques specialised for genotyping particular variant types (see Supplementary Materials 2; Table S1). Although analytical accuracy will vary across the genome and other sources of error exist, these heuristics are sufficient for our purposes.

Table 1 presents the parameter estimates and post-test disease risks calculated across various scenarios. A comprehensive description of parameter ascertainment, including penetrance estimation, is given in Supplementary Materials 3; Table S2 summarises the assumptions and corresponding reality.

Case 1 – Huntington’s disease

HD is a late-onset Mendelian disease with autosomal dominant inheritance caused by a trinucleotide, CAG, short tandem repeat expansion (STRE) in the \(HTT\) gene (OMIM: 613004). We let \(M\) be a CAG expansion of >40 repeat units, which would have complete penetrance in a normal lifespan (29).
We modelled two scenarios for this example, (1) as in genetic screening, defining pre-test disease probability by baseline risk of HD in a general population and (2) as a targeted test, considering pre-test disease probability for an individual whose parent harbours the fully penetrant HTT STRE and who has a 0.5 probability of inheriting M (we have not modelled genetic anticipation [35]).

Case 2 – amyotrophic lateral sclerosis
ALS is a late-onset disease with locus and allelic heterogeneity; genetic associations with risk and phenotype modification range between monogenic and polygenic. Variants in at least 40 genes are associated with ALS [30, 36-38]. SOD1 (OMIM: 147450) and C9orf72 (OMIM: 614260) are the most frequently implicated genes, where variants of each account for fewer than 10% of cases. Autosomal dominant inheritance is typical for most people with a known genetic disease cause.

We modelled several definitions of markers for ALS risk, drawing from three of the commonest ALS genes: SOD1, C9orf72, and FUS (OMIM: 137070). SOD1- and FUS-linked ALS is typically attributed to SNVs and many pathogenic variants with varying strength of supporting evidence have been reported in these genes [39]. The pathogenic form of C9orf72 is a hexanucleotide, GGGGCC, STRE associated principally with the onset of either one or both of ALS and frontotemporal dementia [40]. Known variants in these genes have typically incomplete penetrance; examples include ~90-100% penetrance for SOD1 p.A5V and ~45% for the C9orf72 STRE [41, 42].

The definitions of M modelled in this case study were:

- **SOD1 (all)** – M includes any rare variant reported in people with ALS of European ancestry contained within the meta-analysis sample set from which the variant frequencies were derived (see Supplementary Materials 3.2) [37].
- **SOD1 (A5V)** – M represents the pathogenic SOD1 p.A5V variant, one of the most common SOD1 variants among North American ALS populations, characterised by high penetrance [41, 43].
- **FUS (all)** – M includes any rare variant reported in people with ALS of European ancestry contained within the meta-analysis sample set from which the variant frequencies were derived (see Supplementary Materials 3.2) [37].
- **FUS (ClinVar)** – M includes any of 21 FUS variants reported as pathogenic or likely pathogenic for ALS within ClinVar and present within databases of familial and sporadic ALS (see Table S3) [32-34].
- **C9orf72** – M represents a pathogenic C9orf72 STRE of 30≤ repeat GGGGCC units within the first intron of the C9orf72 gene.

In the SOD1 (all) and FUS (all) scenarios, M encompasses variants classed as pathogenic, likely pathogenic, and variants of uncertain significance [44]. It is appropriate to model these scenarios since many variants of uncertain significance in ALS-implicated genes have a high probability of being deleterious and should not necessarily be ignored [38].

For the C9orf72 marker, we modelled two scenarios: (1) genetic screening with sensitivity and specificity defined by performance of existing tools for genotyping STREs in sequencing data; (2) using repeat-primed polymerase chain reaction with amplicon-length analysis [45] as a secondary test to validate a positive result from screening via sequencing in scenario 1.
PKU is an autosomal recessive disease with infantile onset, caused by variants in the PAH gene (OMIM: 612349) (31); most variants pathogenic for PKU are SNVs. M represents being homozygous or compound heterozygous for any of the three most common PAH variants recorded in European populations of people with PKU: p.Arg408Trp, c.1066-11G>A, p.Arg261Gln (31).

We modelled two testing scenarios for PKU: (1) genetic screening, with pre-test disease probability defined per the baseline population risk of PKU, and (2) secondary testing to confirm positive results obtained using tandem mass spectrometry (46) as in established metabolic screening protocols (see Supplementary Materials 3.3).
Table 1. Input parameters and disease risk estimates following testing for all case study scenarios.

Parameter ascertainment is comprehensively described within Supplementary Materials 3. HD = Huntington’s disease; ALS = amyotrophic lateral sclerosis; PKU = phenylketonuria; SNV = single nucleotide variant; STRE = short tandem repeat expansion. Estimates are based on populations of predominantly European ancestry – 95% confidence intervals shown for newly derived estimates in the ALS case study; includes FUS variants classified as pathogenic or likely pathogenic for ALS in the ClinVar database and recorded within ALS population databases (see Table S3, Table S4). Defined by variant calling performance benchmarks of tools for genotyping in sequencing data by variant type (see Table S1) and, where marked †, by aggregate laboratory accuracy for genotyping C9orf72 STRE with repeat-primed polymerase chain reaction and amplicon-length analysis (45). Risk following positive results in primary screening and confirmatory tests relative to a negative screening result (probability of PKU given a negative metabolic screening result is approximated as $1 \times 10^{-6}$).

<table>
<thead>
<tr>
<th>Case study</th>
<th>Gene containing marker (case study scenario)</th>
<th>Variant type</th>
<th>Pre-test disease probability§</th>
<th>Marker frequency in people affected§</th>
<th>Penetrance§</th>
<th>Test sensitivity¶</th>
<th>Test specificity¶</th>
<th>Disease risk after positive test</th>
<th>Disease risk after negative test</th>
<th>Relative disease risk after positive rather than negative test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: HD</td>
<td>HIT (screening)</td>
<td>STRE</td>
<td>0.000410</td>
<td>1.000</td>
<td>1.000</td>
<td>0.990</td>
<td>0.900</td>
<td>0.00404</td>
<td>0.00000456</td>
<td>887</td>
</tr>
<tr>
<td></td>
<td>HTT (targeted)</td>
<td>STRE</td>
<td>0.500</td>
<td>1.000</td>
<td>1.000</td>
<td>0.990</td>
<td>0.900</td>
<td>0.011</td>
<td>82.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SOD1 (all)</td>
<td>SNV</td>
<td>0.00333</td>
<td>0.0188 (0.0138, 0.0238)</td>
<td>0.701 (0.491, 0.926)</td>
<td>0.9996</td>
<td>0.9995</td>
<td>0.109</td>
<td>0.00327</td>
<td>33.3</td>
</tr>
<tr>
<td></td>
<td>SOD1 (A5V)</td>
<td>SNV</td>
<td>0.00333</td>
<td>0.000529 (4.43x10^-5, 0.00101)</td>
<td>0.91</td>
<td>0.9996</td>
<td>0.9995</td>
<td>0.00683</td>
<td>0.00333</td>
<td>2.05</td>
</tr>
<tr>
<td></td>
<td>FUS (all)</td>
<td>SNV</td>
<td>0.00333</td>
<td>0.00425 (0.0023, 0.0062)</td>
<td>0.579 (0.291, 0.884)</td>
<td>0.9996</td>
<td>0.9995</td>
<td>0.0302</td>
<td>0.00332</td>
<td>9.09</td>
</tr>
<tr>
<td></td>
<td>FUS (ClinVar*)</td>
<td>SNV</td>
<td>0.00333</td>
<td>0.00251 (0.00125, 0.00377)</td>
<td>0.536 (0.211, 0.877)</td>
<td>0.9996</td>
<td>0.9995</td>
<td>0.0194</td>
<td>0.00333</td>
<td>5.84</td>
</tr>
<tr>
<td></td>
<td>C9orf72</td>
<td>STRE</td>
<td>0.00333</td>
<td>0.0635 (0.0538, 0.0732)</td>
<td>0.439 (0.358, 0.520)</td>
<td>0.990</td>
<td>0.900</td>
<td>0.00519</td>
<td>0.00313</td>
<td>1.66</td>
</tr>
<tr>
<td></td>
<td>C9orf72 (positive screening screening confirmation)</td>
<td>STRE</td>
<td>0.0052</td>
<td>0.0635 (0.0538, 0.0732)</td>
<td>0.439 (0.358, 0.520)</td>
<td>0.95†</td>
<td>0.98†</td>
<td>0.0198</td>
<td>0.00489</td>
<td>4.06 (6.35 †)</td>
</tr>
<tr>
<td>2: ALS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3: PKU</td>
<td>PAH (screening)</td>
<td>SNV</td>
<td>0.000100</td>
<td>0.743</td>
<td>0.892</td>
<td>0.9996</td>
<td>0.9995</td>
<td>0.127</td>
<td>0.0000257</td>
<td>4,961</td>
</tr>
<tr>
<td></td>
<td>PAH (positive metabolic screening confirmation)</td>
<td>SNV</td>
<td>0.167</td>
<td>0.743</td>
<td>0.892</td>
<td>0.9996</td>
<td>0.9995</td>
<td>0.889</td>
<td>0.0497</td>
<td>17.9 (889,000 ‡)</td>
</tr>
</tbody>
</table>
Results and Discussion

Post-test disease probability

Screening versus diagnostic testing

Across case studies, we showed low probability of disease following positive results within contextually-blind genetic screening scenarios; risk ranged between 12.7% and 0.4% (see Table 1). Disease risk was always negligible following a negative test result, indicating absence of the tested variant, as would be seen for any rare trait.

The HD case study illustrates the distinction between contextually-blind screening and diagnostic testing for rare diseases: following a positive test result, lifetime HD risk was high (90.8%) using targeted testing but low (0.4%) in screening. This difference reflects that, unlike screening, targeted testing is performed based on some indication of a person’s elevated disease risk (e.g., existing disease symptoms or family history). Accordingly, pre-test disease probability is greater. Inherently low pre-test disease probability will be a pervasive issue in screening for rare diseases.

Relative risk and secondary testing

The utility of a test for identifying at-risk individuals can be examined based on relative disease risk following positive rather than negative test results: utility is limited when risk is only minimally greater for people testing positive rather than negative. This is observed in the ALS C9orf72 case study, where risk was only 1.7 times greater (Table 1) following a positive screening from sequencing alone, despite this variant being the most common genetic cause of ALS (37).

We additionally observed 6.35 times greater ALS risk for a person testing positive on both screening for C9orf72 and a secondary test than for a person testing negative on the initial screening. This increased relative risk reflects that a person testing positive on two independent measures of disease risk has greater absolute probability of disease than after the initial screening result alone.

We emphasise that secondary testing is important to increase certainty in positive tests. The PKU case study demonstrates its potentially sizable impact. A positive screening result using the established metabolic approach alone indicated 16.7% PKU risk, versus 12.7% within genetic screening. The metabolic marker, which is universal across people with PKU and indicates existing disease manifestation, eclipses need for genetic screening for PKU, marked by variants with incomplete penetrance that are not present for all people with PKU. However, the genetic test remains useful for validating the positive metabolic screening result (17): probability of PKU following a confirmatory genetic test conducted on the basis of a positive metabolic screening result was 88.9%.

The overall benefit of secondary testing will however differ by scenario. For ALS, the risk remained moderate (~2%) despite two positive test results for C9orf72.

Constraints upon post-test disease probability

Figure 1, Figure 2, and Figure 3 illustrate how the post-test disease probability reduces as the probability of any test, disease, or marker characteristic decreases. Sensitivity and
Specificity critically constrain certainty about post-test disease risk, and this role is amplified as the other parameter probabilities decrease. Figure 1 particularly demonstrates the increased role of specificity in rarer diseases, where disease risk following a positive test result will be moderated only by penetrance in a protocol of perfect specificity.

Figure 1. Probability of a disease given a positive genetic test result for a marker of increased disease risk ($P(D|T)$) according to the sensitivity ($P(T|M)$) and the specificity ($P(T'|M')$) of the testing protocol.

Panel A presents this for a disease with pre-test probability ($P(D)$) of 0.5, while panel B presents a disease with $P(D)$ of 0.01. Penetrance is complete ($P(D|M) = 1$) and variant M is harboured by all people with disease D ($P(M|D) = 1$) in both panels.
Figure 2. Probability of disease D following a positive genetic test result for marker M (P(D|T)) according to pre-test disease probability (P(D))

$M$ occurs in all people with $D$ ($P(M|D) = 1$) and has complete penetrance ($P(D|M) = 1$). Plot lines are defined according to sensitivity ($P(T|M)$) and specificity ($P(T'|M')$) of existing protocols for genotyping variant types (see Supplementary Materials 2): single nucleotide variant (SNV), $P(T|M) = 0.9996$; small insertion or deletion (Indel), $P(T|M) = 0.9962$; short tandem repeat expansion (STRE), $P(T|M) = 0.99$; copy number variant (CNV) – del. (deletion), $P(T|M) = 0.289$, $P(T'|M') = 0.959$; CNV – dup. (duplication), $P(T|M) = 0.1020$, $P(T'|M') = 0.9233$. 
Figure 3. Change in disease risk following a positive test result for a marker of increased disease risk (P(D|T)) according to penetrance (P(D|M)).

Panels A and B display modelled and hypothetical markers of ALS which differ in frequency across people affected by ALS (P(M|D)), where pre-test disease probability (P(D)) is 0.0033 and diamonds mark the penetrance estimated for non-hypothetical variants (see Table 1). Panels C and D display diseases in which P(M|D) = 1 and with P(D) set in line with the modelled case studies or a hypothetical rare disease. Analytic validity parameters are defined according to the performance of sequencing tools for genotyping single nucleotide variants in A and C, and of short tandem repeat expansions in B and D (see Table S1; Figure 2).
As is well-recognised, high sensitivity and specificity are essential to maximise utility of testing. The respective trade-offs between prioritising each of these must be regarded: high sensitivity is required to detect markers, while high specificity increases confidence in positive results. Established screening protocols prioritise high sensitivity to maximise detection of at-risk individuals, with confirmatory secondary testing being vital to minimise false-positive results (9, 22, 23).

Since the characteristics of diseases and associated variants are all pre-determined within a population, disease markers (variants) screened should be chosen carefully. The most useful variants will be those more prevalent among people affected, of high penetrance, and which can be genotyped with high sensitivity and specificity (see Table 1, ALS case study; Figure 3, panels A and B).

Practical implementation of genetic screening

Marker selection

Before a marker is used in screening, its relevance across people must be evaluated, recognising that this may vary by ancestry. For instance, particular variants may be less common or only present in certain populations, and penetrance can also vary between them (36, 43). Screening protocols must therefore account for these differences to prevent systemic inequalities, especially for minorities which are often under-studied and therefore have limited genetic information available.

Regard must be given to the clinical interpretability of selected markers. We illustrated several approaches to defining markers in the ALS case study. Within the SOD1 (all) scenario, disease risk is marked by an aggregation of putatively pathogenic SOD1 variants. Without curation, the relationship to disease varies across them. E.g., SOD1 p.I114T has substantially lower penetrance than p.A5V, and many potentially relevant variants have uncertain significance (34, 38, 39, 42). Curation could involve defining a positive result as presence of variants with sufficient evidence of pathogenicity (44), as in the FUS (ClinVar) scenario, or as harbouring specific variants, as in the SOD1 (A5V) scenario.

De novo variants and variants of uncertain significance present a substantial challenge for screening since they will frequently be identified, yet must be set aside until variant interpretation is possible despite potentially being deleterious (12). PKU demonstrates the scale of this issue for rare diseases with multiple implicated variants, as 55% of deleterious PAH genotypes are observed uniquely (31).

Utility over time and actionability

As genetic screening is possible from birth, while non-genetic methods may not be, age of viability for screening methods should be evaluated. For late-onset diseases, early genetic screening may enable preventative treatments to at-risk individuals or close monitoring for prodromal disease markers. For instance, rapid eye movement sleep behaviour disorder is a prodromal feature for Parkinson’s disease (47) and monitoring of at-risk individuals may enable early intervention. The influence of time upon treatment viability and effectiveness must also be considered. For example, genetic therapy has potential utility for preventing or delaying onset of degenerative disorders (48).
The ultimate benefit of early identification of disease risk through genetic screening is contingent upon the actionability of the result. A framework of actionability (24, 25), shown to align with laypersons’ views on treatment acceptability (49), states that actionability is determined by: disease likelihood and severity, intervention effectiveness in disease minimisation or prevention, and the consequence of the intervention to a person and risk if not performed. Each of these elements are critical when selecting traits and markers for genetic screening and for the clinical interpretation of results.

Limitations

The Bayesian logic in the case studies simplifies genotype-phenotype relationships and cannot address all considerations relevant to clinical genetic testing. Phenotype variability is not considered. Other factors include: polygenicity and oligogenicity, pleiotropy, the role of genetic and environmental modifying factors, and that of additive genetic effects in recessive conditions and heterozygous carriers of pathogenic variants. Such influences can fundamentally impact both the probability that a disease will manifest and its severity. For instance, although spinal muscular atrophy is caused by partial or complete biallelic deletion of the SMN1 gene, additional copies of SMN2 reduce disease expressivity by mitigating loss of SMN1 function (50). Any results from genetic screening must be interpreted within the wider context of that disease and its modifiers.

Conclusion

We have shown that risk following a positive screening test result can be strikingly low for rare neurological diseases. Accordingly, to maximise the utility of screening, we suggest prioritising protocols of very high sensitivity and specificity, careful selection of markers for screening, giving regard to clinical interpretability, and secondary testing to confirm positive findings.

A key advantage of a genetic screening approach for late-onset diseases is that these markers can be examined across the lifespan. Hence, positive test results could be useful for targeting people for prevention, and for monitoring of prodromal features.

Although considering disease risk within a Bayesian context is not novel, it is important to stress the considerations raised here at a time when governments evaluate implementation of genomic sequencing for population screening and as access to genetic testing outside healthcare settings increases. While genetic screening has many potential benefits, the limitations of such an approach should be properly understood. Policy makers must consider the impact of a positive test result on large numbers of people that will never develop a given disease, a particularly salient issue for late-onset diseases. Although not the present focus, the substantial ethical and social considerations raised in conjunction to screening must also be regarded (13).

Data Availability

Data sharing not applicable
Acknowledgements

We wish to thank Professor Naomi Wray of the Queensland Brain Institute for her contribution to project conceptualisation and methodology. We also thank Dr Leslie Biesecker of the National Human Genome Research Institute of the National Institutes of Health for contributing to our discussions around this work.

Author contributions


Funding

AAC is an NIHR Senior Investigator. This is an EU Joint Programme-Neurodegenerative Disease Research (JPND) project. The project is supported through the following funding organizations under the aegis of JPND (United Kingdom, Medical Research Council MR/L501529/1 and MR/R024804/1; Economic and Social Research Council ES/L008238/1) and through the Motor Neurone Disease Association. This study represents independent research part funded by the NIHR Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King’s College London. The work leading up to this publication was funded by the European Community’s Horizon 2020 Programme (H2020-PHC-2014-two-stage; grant 633413). This work was part funded by my Name’s Doddie Foundation, and Alan Davidson Foundation. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, King’s College London, or the Department of Health and Social Care. AI is funded by South London and Maudsley NHS Foundation Trust, MND Scotland, Motor Neurone Disease Association, National Institute for Health and Care Research, Spastic Paraplegia Foundation, Rosetrees Trust, Darby Rimmer MND Foundation, the Medical Research Council (UKRI) and Alzheimer’s Research UK.

Role of the funding source

The funding sources had no role in the writing of this manuscript or the decision to submit it for publication.

Declaration of interests

AI, MR, FF, and NP declare no competing interests. TPS declares employment with AstraZeneca. AAC declares royalties from The Brain, A Beginner’s Guide: One World Publications, and The Genetics of Complex Human Diseases, A Laboratory Manual. AAC reports consultancies or advisory boards for Amylyx, Apellis, Biogen, Brainstorm, Cytokinetics, GenieUs, GSK, Lilly, Mitsubishi Tanabe Pharma, Novartis, OrionPharma, Quralis, Sano, Sanofi, and Wave Pharmaceuticals. AAC is the deputy editor for Brain. AAC has planned patents for ‘Use of CSF-Neurofilament determinations and CSF-Neurofilament thresholds of prognostic and stratification value with regards to response to therapy in neuromuscular and neurodegenerative diseases’, AAC is Scientific co-lead for United to End MND Partnership Coalition and Guarantor of Brain. AAC has stock options for Quralis.
References:


33. ALS Variant Server [Internet]. [cited 02/2021]. Available from: http://als.umassmed.edu/.


