Design and Reporting Considerations for Genetic Screening Tests

Jill Hagenkord, MD, FCAP, MDisrupt

Birgit Funke, PhD, FACMG, Veritas Genetics and Harvard Medical School

Emily Qian, CGC, Veritas Genetics

Madhuri Hegde, PhD, FACMG, Perkin Elmer

Kevin B Jacobs, Progenity Inc.

Matthew Ferber, PhD, FACMG, Mayo Center for Individualized Medicine and Mayo Medical Laboratories

Matthew Lebo, PhD, FACMG, Laboratory for Molecular Medicine, Partners Personalized Medicine and Harvard Medical School

Adam, Buchanan, CGC, Geisinger Health System

David Bick, MD, Hudson Alpha Institute for Biotechnology

Short Running Head: Design and Reporting Considerations for Genetic Screening Tests

Number of Text Pages: 12, number of Tables: 3, number of Figures: 2

Corresponding Author: jill@mdisrupt.com, 650-815-6620, 3031 Tisch Way, San Jose, CA 95128

Key words: genetic screening, epidemiologic methods, clinical validity, positive predictive value, clinical test validation

Conflict of Interest Statement

Dr. Hagenkord is a shareholder in Color Genomics. Dr. Funke and Ms. Qian are shareholders in and/or employees of Veritas Genetics. Dr. Hegde is a shareholder in and employee of Perkin Elmer. Mr. Jacobs is a shareholder in an employee of Progenity. The remaining authors have no conflict of interest to declare.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

ABSTRACT

Testing asymptomatic individuals for unsuspected conditions is not new to the medical and public health communities and protocols to develop screening tests are well-established. However, the application of screening principles to inherited diseases presents unique challenges. Unlike most screening tests, the natural history and disease prevalence of most rare inherited diseases in an unselected population are unknown. It is difficult or impossible to obtain a “truth set” cohort for clinical validation studies. As a result, it is not possible to accurately calculate clinical positive and negative predictive values for “likely pathogenic” genetic variants, which are commonly returned in genetic screening assays. In addition, many of the genetic conditions included in screening panels do not have clinical confirmatory tests. All of these elements are typically required to justify the development of a screening test, according to the World Health Organization screening principles. Nevertheless, as the cost of DNA sequencing continues to fall, more individuals are opting to undergo genomic testing in the absence of a clinical indication. Despite the challenges, reasonable estimates can be deduced and used to inform test design strategies. Here, we review test design principles and apply them to genetic screening.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

Introduction

The design and validation of genetic tests has unique challenges when compared to other types of clinical tests. In genetics, many conditions do not have a “truth set” cohort for use in the clinical validation study to establish the clinical sensitivity, specificity, and predictive values of the test. In addition, the natural history of many conditions and associated epidemiology may be incompletely understood. To overcome these limitations, genetics professional societies have utilized reasonable estimates to develop test design, variant classification, and reporting guidelines when testing individuals suspected of having an inherited condition (i.e., “diagnostic” testing). DNA sequence-based inherited genetic testing has been conducted primarily in patients with a suspected genetic condition, and the existing guidelines were optimized for this specific intended use. We are now entering an era in which DNA sequencing is carried out in different populations with different intended uses (Figure 1A). Application of the current guidelines to a test with a different intended use could result in an unacceptable number of false results.

Unlike diagnostic testing, screening tests are intended to identify the presence of an as-yet-undiagnosed condition in individuals without signs or symptoms.[1,2] Screening tests provide an opportunity for early detection and/or prevention. Compared to diagnostic tests, however, screening tests have a higher risk of false positives and the need to mitigate the possible clinical impact of false positive results. The principles of screening and test design strategy are well-established in the medical literature.[3] First published in 1968 by the World Health Organization[4] they were adapted to DNA-based preventive screening in 2003.[5] Candidate conditions for screening should have a well-understood natural history, established prevalence, a clinical confirmatory test to identify false positives, as well as safe, effective, and accessible preventive actions. The test itself must have well-defined performance metrics, including predictive values, and must be optimized for individuals with a low prior probability of disease.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

Genetic screening tests are now offered by several laboratories in several different contexts, such as the issuance of secondary reports in the context of a diagnostic exome (opportunistic screening), elective sequencing paid for by a curious individual, population screening offered by health systems and employers, and research studies involving return of results to participants.[6,7] The panel of genetic conditions included in genetic screens vary by laboratory, but commonly offered panels include the CDC Tier 1 conditions and the ACMG Secondary Findings V2.0 (ie, ACMG59^{TM}). The CDC Tier 1 conditions are hereditary breast and ovarian cancer syndrome (HBOC), Lynch Syndrome (LS), and Familial Hypercholesterolemia (FH).[8] The ACMG has issued recommendations for reporting secondary findings in clinical exome and genome sequencing. They enumerate 27 conditions (59 genes) for inclusion in an opportunistic screen.[9] Larger genetic screening panels are commercially available.

When evaluating whether a condition is appropriate to include in a screening test, the WHO criteria require that the natural history and prevalence of the condition in the test population are well-understood and that safe, effective, preventive options are available. The CDC Tier 1 conditions have been studied in large unselected populations, the disease prevalences (and penetrance) are published, and preventive guidelines appropriate for the test population are available. However, for the ACMG59^{TM} conditions, the prevalence and penetrance in an unselected population are mostly unknown. In fact, recent studies have shown that several of these conditions appear to have a very low penetrance in an unselected population.[10,11] The ACMG has recently re-emphasized that this set of genes in not appropriate for general screening until their natural history and epidemiology are better understood.[12]

In addition, many conditions commonly included in genetic screening tests do not have a reference standard test/confirmatory test. In the context of screening, a confirmatory test adjudicates the presence or absence of a medical *condition* and should not be confused with an orthogonal analytical confirmation that confirms the presence or absence of a genetic *variant*. Examples of screening tests and clinical confirmatory tests include mammogram followed by

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

biopsy and non-invasive prenatal testing (NIPT) followed by karyotype. (Table S1) As reviewed below, the lack of a clinical confirmatory test makes it difficult or impossible to obtain a "truth set" cohort needed to define the clinical accuracy of the test. The lack of a clinical confirmatory test also results in the inability to exclude false positives with a confirmatory test, as is typical for most screening tests.

The inconsistent application of screening and test design principles across genetic laboratories creates uncertainty concerning the clinical implications of a positive result and thus the appropriateness of providing or paying for preventive interventions. It increases the possibility that the same variant can be interpreted as positive (and actionable) by some laboratories, but negative by others. It is critical that physicians are aware of the predictive value of a positive screening test result and the presence or absence of a confirmatory test. This is especially true in delivery models where there is a separation between the ordering physician, the patient, and the treating physician, such as when a commercial lab uses an independent third-party network of physicians to order the test for the customer online.

Review of Test Design Principles

Methods for clinical test design are well-established and covered in depth in the literature.[13,14] In development, the new test is performed on a cohort of samples from individuals known to have the condition and known to be free of the condition. The disease state of each of these samples is established by an existing reference standard test, and this cohort serves as the "truth set" for validation studies. The agreement between the new test (index test) and the "truth set" determines the clinical sensitivity and specificity.[1] Ideally, the new test will discriminate between these cohorts perfectly; but in practice there will be false positive and false negative results. Sensitivity and specificity measure the test's ability to accurately identify the presence or absence of disease *in a cohort where disease presence or absence is already known*. (Table 1)

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

The predictive values provide the post-test probability of disease *for an individual*.[15] Predictive values address the probability that a person with a positive result actually has the condition (positive predictive value or PPV) or that the person with a negative result does not have the condition (negative predictive value or NPV).[2] In the clinic, the PPV and NPV are more useful than sensitivity and specificity since in practice the presence/absence of disease is unknown prior to testing.[16] The predictive values also help inform the most appropriate cut-off threshold between a positive result and negative result as appropriate for the test's intended use. Sliding the cut-off towards higher specificity will result in fewer false positives, while sliding the cut-off toward higher sensitivity will decrease false negatives.[17] (Figure 1) When designing a screening test, the extent to which true positive and true negative results are medically desirable and the extent to which false positive and false negative results are tolerable or even acceptable must be weighed.[2] Considerations include the immediate and long-term burden on the healthcare system, the treatability of the condition, psychosocial effects, and the potential over-utilization of invasive procedures, diagnostic procedures, or surveillance.[1,2]

Unlike sensitivity and specificity, PPV and NPV will vary depending on the prevalence of the disease in the test population. For tests with very high sensitivity and specificity, the PPV may still be quite low, if the disease prevalence is low. At very low prevalence, small changes in specificity can dramatically reduce the PPV. (Figure 2). It is noteworthy that diminishing disease prevalence in a test population adversely impacts PPV more than NPV. When disease prevalence is very low, increasing the specificity will improve the PPV, but have negligible impact on the sensitivity (and NPV).[17,18] Since sensitivity has a greater influence on NPV and specificity has a greater influence on PPV, screening test discussions focus primarily on PPV and specificity.

There are three test design options that will increase the PPV of a screening test: 1) increase the specificity of the screening test, 2) pair the screening test with a clinical confirmatory test, and 3) employ the screening test in a population with a higher disease prevalence. In genetic

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

screening of unselected populations, increasing the specificity of the genetic screening test itself is the most viable option. It is described in detail below.

Applying Screening Principles to Genetics

The test design terminology used widely in laboratory medicine is not always consistently applied to genetic testing. Table 1 provides general definitions of key assay development terms and examples of these terms applied to genetics. A potential source of confusion stems from the fact that PPV and penetrance both address the probability of phenotypic manifestation, but they address very different probabilities. PPV applies to all types of tests, it is dependent primarily on the test's specificity and disease prevalence in the test population, and it provides information about false positives. In the context of genetics, PPV addresses whether a patient with a positive result has the associated inherited condition, such as HBOC. PPV does not address the penetrance of the condition, the types of cancers that are associated with a given hereditary cancer syndrome, nor the patient's risk of developing one of those cancers. In contrast, penetrance is a feature unique to genetic conditions (not tests) and is only applicable once a positive result is confirmed to be a true positive. Penetrance addresses the chance that a person diagnosed with a genetic condition, such as HBOC, will develop certain kinds of cancers. Uncertain penetrance is another unique challenge in genetic testing.[8,19,20]

As mentioned, a significant challenge of genetic test design is the lack of a large "truth set" cohort, established by a reference standard test, that can be used for clinical validation. Ideally, during test development, a large cohort of samples from individuals known to have a genetic condition and large cohort known to be free of the condition would be used to determine clinical sensitivity and specificity of the genetic test. This "truth set" would enable the developer to determine the number of false positive results at different levels of specificity, and then set a cut-off at the desired level. In the absence of a "truth set cohort", clinical validity is estimated from the level of confidence that a specific variant does or does not cause disease. The minimum certainty that a variant causes disease is analogous to the clinical specificity.[18]

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

When testing high-risk individuals, the ACMG/AMP variant classification guideline indicates that an appropriate positive/negative cut-off is when there is at least 90% confidence that a variant is disease-causing. This places the cut-off between variant of uncertain significance (VUS) and likely pathogenic (LP) and triggers the issuance of a positive report when a variant is classified as Pathogenic (P) or Likely Pathogenic (LP) .[21] (Figure 1B). Assuming the confidence in pathogenic variants is 100%, the possibility of clinical false positives is reflected in the minimum confidence level of LPs. “Likely pathogenic” implies that there is less than 100% certainty that the variant causes disease. Some LPs will be false positives. If the minimum confidence is set at 90% for LPs, then 1 in 10 LP variants may be false positives, which is a medically appropriate cut-off when testing high-risk patients. However, the guideline warns that applying this same cut-off to an asymptomatic test population may result in an unacceptable number of false positives given that the disease prevalence is much lower in an unselected population.[21]

Many genetic conditions commonly included in screening panels do not have a clinical confirmatory test. This is especially concerning for conditions where the clinical impact of a false positive is harmful, expensive, or irreversible. Outside of genetics, screening tests are intentionally designed to permit some false positives to avoid missing true positives. These false positives are tolerated because it is standard practice to follow a positive screening test with a more specific confirmatory diagnostic test.[1] For genetic tests that do not have a clinical confirmatory test, this two-step process is not possible, and therefore the specificity of the screening test itself must be high.

Lastly, in the absence of prevalence, penetrance, and predictive values it is impossible to determine the health economic benefit of genetic screening. This may impact the willingness of payers to cover downstream preventive interventions. The cost of managing the false positives may outweigh the potential benefits of identifying the true positives. Simply recommending a lifetime of increased cancer surveillance may seem benign, but it has both cost and risk. The

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

more mammograms a woman undergoes, the more likely she will have a false positive result. This will result in unnecessary anxiety and invasive follow-up tests. The chance of a false positive result after one mammogram is 10%. Younger women are more likely to have a false positive result than older women. After 10 yearly mammograms, the chance of having a false positive is 50-60%.[22] Without data to support the safety and efficacy of an intervention, increased screening followed by increased surveillance may do more harm than good. There are many published examples where interventions assumed to have benefit resulted in considerable harm.[1,2]

In spite of these challenges, we propose that reasonable approximations of specificity, prevalence, and PPV are possible and can be used to guide the development strategy of genetic tests. Consistent application of screening principles and appropriate supporting content in test reports will reduce the potential for over- or under-treatment by the clinician receiving the results.

Estimating Specificity, Prevalence, and PPV in Genetic Screening

We estimated specificity, prevalence and PPV for autosomal dominant monogenic disorders in asymptomatic individuals in an unselected (low risk) population. The approximate specificity and PPV over a range of disease prevalence are presented in Tables 2 and 3. They are based on the following assumptions:

- 100% analytical test sensitivity.
- 100% confidence in a “pathogenic” classification.
- 90% confidence in a “likely pathogenic” classification. In practice, the actual confidence will vary by variant type and gene/condition.
- ⅓ of all positive results are “likely pathogenic” variants and ⅔ of all positive results are “pathogenic”. Together, these compose the “overall positive rate.” In practice, this will vary by laboratory and condition.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

- Prevalence of the CDC Tier 1 conditions range from 1/500 to 1/200 in an unselected population.[8]
- The published range for overall positive rates of ACMG59™ is between 1% and 8.5%. [23–26] We used 3% and 6% as examples in Table 2. The published overall positive rate CDC Tier 1 is approximately 1.5% [27,28]
- Prevalence of ACMG59™ and other inherited diseases in an unselected population is not known but reasonable estimates range from 1/25,000 for the rarest conditions and 1/500 for the more common conditions.[29–32] Several conditions are so rare that firm epidemiological estimates are not available.

The specificity of the test depends on the overall positive rate of the test, the type of variant, the strength of the gene-disease association, and knowledge of the specific gene/disorder. The specificity provided during assay development may not be representative of the test population in terms of disease prevalence. For example, the test cohort may have been comprised of an equal number of samples from patients with disease and patient free of the disease. However, the overall positive rate *in practice* is not 50%, and therefore the clinical specificity will need to be recalculated. Supplemental Figure 1 provides an example of how the overall positive rate is used to determine the specificities found in Tables 2 and 3.

Similar to other authors, the calculations in Tables 2 and 3 used the assumption that all LP variants have a confidence level of 90%.[18,21,33] In the absence of empirical data obtained from comparison with a reference standard test, 90% is admittedly an *estimate* of the minimum confidence. In practice, the level of confidence will vary depending on the variant. For a monogenic disorder with a dominant inheritance pattern caused primarily by loss of gene function from simple variants, the specificity of variant classification will generally decrease for each of the following functional categories (highest specificity to lowest specificity): known pathogenic variants, loss-of-function variants in relevant gene domains (nonsense and frameshift variants), canonical splice altering variants and non-canonical splice variants with

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

some functional evidence, missense variants with credible functional evidence. This trend toward decreasing specificity is due to the inherent complexity of interpretation and the rate of errors that can occur in applying evidence toward a classification. Specificity (confidence) will be 100% for well-known pathogenic variants. This will decrease to below 90% in categories near the end of the list. As with other types of clinical tests, the positive rate will increase as more genes and variant types are tested or when less stringent criteria are used to specify a variant as "positive." The false positive rate will increase as the overall positive rate increases.

In practice, the PPV for likely pathogenic variants should be estimated for each condition included in a genetic screening panel. The estimates for commonly used screening panels provided in Tables 2 and 3 reveal informative trends and underscores the need for consistent understanding and application of test design principles. Two possible overall positive rates are shown for ACMG59™ conditions in order to demonstrate that higher overall positive rates will lower the PPV. (Table 2). Note that specificity decreases as the number of LPs increase and as panel size increases and, as noted above, small changes in specificity can have a significant impact on PPV when testing for diseases with a low prevalence. Importantly, most conditions included in genetic screening panels have a very low prevalence in an unselected population, and thus may have a corresponding low PPV for LP variants if specificity is not adjusted.

When the prevalence is very low, the false positive rate for a test is greater than the prevalence of the condition. This is known as the False Positive Paradox.[34,35] Under these conditions, more disease-free individuals will test positive than affected individuals. If the ACMG59™ screening test has a 99.9% specificity then the test will find more false positives than true positives for those conditions with a prevalence of 1/1,000 or less.[36] Conditions with a disease prevalence of 1/25,000 will always generate more false positive LPs than true positives, since it would require a specificity of greater than 99.995% to obtain one true positive for each false positive. [18 37]

Recommendations for Genetic Screening Test Design and Reporting

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

As an alternative to estimates of PPV, we considered whether PPV could be extrapolated from intra-laboratory reclassification rates, published reclassification rates from high volume laboratories, or the reclassification rate in public variant databases. We concluded that none of these are a replacement for condition-specific performance metrics derived from comparison with a clinical "truth set." Datasets from high volume laboratories indicate that reclassification rates are laboratory dependent, based on high-risk populations, and fluctuate with differences in test volume in each year, the rate at which new data appears that affect classes of variants, and how many genes were available for testing each year.[38,39] Although there are some genetic conditions that have a potential reference standard test, such as coagulopathies confirmed by mixing studies, full gene sequencing studies in an unselected population have not been published, and data from coagulopathies could not be applied to other conditions. At present, laboratories offering genetic screening tests will need to use reasonable estimates of PPV for test design and optimization.

Screening tests for inherited genetic disease should adhere to the long-standing principles of screening applied to other types of medical tests. The epidemiology and natural history of the condition should be adequately understood. Screening test performance should be appropriate for the purpose, with all key components of test accuracy established (e.g., sensitivity, specificity and positive predictive value). For screening tests with less than 100% specificity, the presence or absence of a confirmatory test should be considered. Preventive guidelines should exist that have been deemed safe, effective, and cost-effective for the test population. Interventions should be accessible and affordable.[3] Currently, only the CDC Tier 1 conditions meet these criteria.

We considered three possible options for increasing the specificity of a genetic screening test. The laboratory could chose to report only known pathogenic variants as a positive result (100% specificity), report known pathogenic variants and high confidence LP variants as positive (high specificity), or titrate clinical specificity based on clinical implications and availability of a

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

confirmatory functional test.[18] For example, Mayo Clinic's GeneGuide™ only reports a limited list of known pathogenic variants as a positive result, despite their ability to detect novel variants.[40] In the MyCode Community Health Initiative, Geisinger filters out lower confidence likely pathogenic variants and only reports pathogenic and high confidence likely pathogenic variants (together called 'expected pathogenic' variants) as positive results.[27] However, several laboratories have not optimized the PPV of LPs in their genetic screening tests. As a result, there is a possibility of receiving a positive result or a negative result for a lower confidence LP variant depending on the laboratory performing the test. This underscores the need for consistent application of screening principles and clear statements of intended use by genetic laboratories.

Comparing HBOC and FH illustrates how titration of clinical specificity by condition might be applied. A very low false positive rate is required for HBOC because risk-reducing surgery is a management option, increased surveillance has adverse consequences and high cost, and no confirmatory tests exist. Thus, only carefully curated pathogenic variants and, perhaps, high confidence LP variants should be reported in a screening context. For FH, inclusion of rare, likely pathogenic missense variants may be acceptable [18] since a positive FH result can be confirmed with an LDL-level and the impact of a false positive is low. In this case, a higher level of uncertainty can be tolerated.

Two different strategies have been proposed to adjust the specificity of a genetic test according to intended use: 1) adjust the cut-off between positive and negative, [6] or 2) adjust the stringency of variant classification. [21] [18,41] We recommend following the protocol used for other types of clinical tests and simply adjust the positive/negative cut-off as appropriate for the intended use. Variant classification itself should remain indifferent to the intended use. Figure 1B shows how setting the cut off between lower and higher confidence LPs can reduce the false positive rate to an acceptable level for a screening test. It is important to realize that the impact of this approach on clinical sensitivity will be negligible given the low prevalence of these conditions. Including

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

lower confidence novel variants that inherently reduce the specificity of the screening test is of little benefit to the patient.[17,18]

Genetic screening reports should be clear and transparent about the test performance metrics and utility. Genetic screening is often initiated outside the typical clinic setting. In many instances the primary care provider receiving the report did not order the test, is not familiar with the nuances of genetics, and is unaware of the possibility of inter-laboratory variability in test design strategies. The following information should be included in the test report, pre-participation informational collateral, consent forms, and physician education material:

- An intended use statement, including whether it is designed for screening, diagnosis, or monitoring; the specific analyte or condition of interest; the target test population; the technology used; and how the results should be used.
- If LPs are included, provide the estimated PPV for the gene/condition pair in question in the test population, and the assumptions behind the estimate.
- For screening tests, indicate the availability or absence of a clinical confirmatory test.
- For screening tests, describe the clinical features of the syndrome and the penetrance of each feature in an unselected population, if known, or state that it is not known.
- Provide a reference to patient management guidelines that have been proven safe and effective in the test population, if available, or state that none exist.
- Participants should be aware that post-test preventive care may not be covered by insurance. Coverage will depend on the level of evidence for the utility of the intervention and the type of insurance. For most commercial payers, the presence of a positive genetic result is sufficient for diagnosis and coverage of the CDC Tier 1 conditions, but the same is not necessarily true for other insurance plans or conditions outside CDC Tier 1.[42]
- Laboratories should consider conducting comprehension testing on these limitations and/or require positive results to be communicated to the patient by a genetic counselor.
- Laboratories reporting LPs as a positive result for conditions that evoke the false positive paradox should reconsider this practice.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

Examples from current practice

The estimated PPVs from Table 3 (specificity in the range of 99.8% and 99.9%) can be used in conjunction with published penetrance, presence/absence of confirmatory testing, and likelihood of coverage for downstream interventions to help determine which conditions are appropriate for inclusion in a genetic screen and what type of supporting information should be included in the test report. The following are examples of conditions included in some genetic screening tests using these principles:

- **HBOC**, **pathogenic variant in *BRCA1*.** The PPV will be 100% in both low-risk and high-risk populations. The risk of developing breast and ovarian cancer in an unselected cohort is known and appears to be only slightly less than that observed in high-risk cohorts.[43–45] The interventions proven safe and effective for high-risk patients can be applied to individuals from an unselected population.[27] The presence of a positive genetic test result alone meets criteria for most commercial insurance coverage policies for the management of HBOC. Original Medicare may not cover preventive care.
- **HBOC**, **likely pathogenic variant in *BRCA1*.** The PPV will be 80-90% in an unselected population. A clinical confirmatory test is not available to determine if this is a true positive. If it is a true positive, the penetrance is similar to that observed in high-risk cohorts. The appropriate action is undetermined.
- **Peutz-Jeghers syndrome (PJS), likely pathogenic variant in *STK11*.** The prevalence of PJS is less than 1/10,000 and the PPV will be under 10% in an unselected population. A clinical confirmatory test is not available to determine if this is a true positive. If it is a true positive, the penetrance has not been studied in an unselected population. A patient cannot be counseled about cancer risk based on this result. Additional surveillance has not been demonstrated to be safe, useful, or cost-effective. Insurance coverage for increased surveillance may be challenged.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

- **Familial Adenomatous Polyposis (FAP), pathogenic variant in *APC*.** The PPV will be 100%. The person has FAP or attenuated FAP. However, the penetrance of polyposis for individuals from an unselected population appears to be very low, although extracolonic features may occur.[11] The clinical guidelines for managing FAP patients is not appropriate since polyposis is unlikely. No guidelines exist detailing the best management for this patient. Increased surveillance has not been demonstrated to be safe, useful or cost-effective. Insurance coverage for increased surveillance may be challenged.
- **Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC), likely pathogenic variant in *DSP*.** The prevalence is approximately 1/5000[46] and the PPV is estimated to be less than 17%. The studies conducted on an unselected population suggest that the penetrance is very low.[10] There is no definitive diagnostic standard. No guidelines exist detailing the best management or counseling for this patient. Increased surveillance has not been demonstrated to be safe, useful, or cost-effective. Insurance coverage for increased surveillance may be challenged. In addition, one of the key risk management recommendations - limiting exercise - runs counter to physical activity recommendations known to be beneficial for a wide range of health conditions.
- **Familial Hypercholesterolemia (FH) likely pathogenic variant in *LDLR*.** The prevalence of FH in an unselected population is approximately 1/220.[28,47,48] The estimated PPV is approximately 90%. A functional study, such as LDL cholesterol level, can help adjudicate whether this is a true positive or a false positive. Since the penetrance of FH has been determined to be approximately the same in high risk and unselected cohorts, it is appropriate to manage confirmed cases according to FH guidelines. [49] The presence of a positive genetic test result alone meets criteria for most commercial insurance coverage policies for the preventive management of FH.

Conclusion

Genetic testing of low-risk individuals is occurring more frequently and is likely to increase significantly in the near future. In the absence of penetrance and prevalence data, with few confirmatory tests, and few clinical utility guidelines describing preventive interventions, some

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

would argue that genetic testing of low-risk individuals should not be conducted or restricted to certain circumstances, such as opportunistic screening in the context of a genetically literate care team. However, due to consumer demand, genetic testing of low-risk individuals will likely proceed. Laboratories can mitigate the risk by modeling estimates of the missing data, by designing tests according to their intended use, by adhering to established principles of screening, and by providing consumers and physicians abundantly clear limitations to the clinical utility of the results.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

References

1. Maxim, L. D., Daniel Maxim, L., Niebo, R. & Utell, M. J. Screening tests: a review with examples. *Inhalation Toxicology* **26**, 811–828 (2014).
2. Trevethan, R. Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. *Front Public Health* **5**, 307 (2017).
3. Dobrow, M. J., Hagens, V., Chafe, R., Sullivan, T. & Rabeneck, L. Consolidated principles for screening based on a systematic review and consensus process. *CMAJ* **190**, E422–E429 (2018).
4. Wilson, J. M. G., Jungner, G. & Health Organization, W. Principles and practice of screening for disease. (1968).
5. McCabe, L. L. & McCabe, E. R. B. Population screening in the age of genomic medicine. *J. Med.* (2003).
6. Lu, J. T. *et al.* Evaluation for Genetic Disorders in the Absence of a Clinical Indication for Testing: Elective Genomic Testing. *J. Mol. Diagn.* **21**, 3–12 (2019).
7. Brothers, K. B., Vassy, J. L. & Green, R. C. Reconciling Opportunistic and Population Screening in Clinical Genomics. *Mayo Clin. Proc.* **94**, 103–109 (2019).
8. Murray, M. F., J. P. Evans, M. Angrist, K. Chan, W. Uhlmann, D. L. Doyle, S. M. Fullerton, T. Ganiats, J. Hagenkord, S. Imhof, S. H. Rim, L. Ortmann, N. Aziz, W. D. Dotson, E. Matloff, K. Young, K. Kaphingst, A. Bradbury, J. Scott, C. Wang, A. Zauber, M. Levine, B. Korf, D. Leonard, C. Wicklund, G. Isham, and M. J. Khoury. A Proposed Approach for Implementing Genomics-Based Screening Programs for Healthy Adults. *NAM Perspectives, National Academy of Medicine, Washington, DC.* (2018). doi:10.31478/201812a
9. Kalia, S. S. *et al.* Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2. 0): a policy statement of the American College of Medical Genetics and Genomics. *Genet. Med.* **19**, 249 (2017).
10. Haggerty, C. M. *et al.* Electronic health record phenotype in subjects with genetic variants associated with arrhythmogenic right ventricular cardiomyopathy: a study of 30,716

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

subjects with exome sequencing. *Genet. Med.* **19**, 1245–1252 (2017).

11. Rocha, H. APC Variant Identification in an Unselected Patient Population: Where are the Polyps?
12. ACMG Board of Directors. The use of ACMG secondary findings recommendations for general population screening: a policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* **21**, 1467–1468 (2019).
13. Jennings, L., Van Deerlin, V. M., Gulley, M. L. & College of American Pathologists Molecular Pathology Resource Committee. Recommended principles and practices for validating clinical molecular pathology tests. *Arch. Pathol. Lab. Med.* **133**, 743–755 (2009).
14. Christenson, R. H. & Committee on Evidence Based Laboratory Medicine of the International Federation for Clinical Chemistry Laboratory Medicine. Evidence-based laboratory medicine - a guide for critical evaluation of in vitro laboratory testing. *Ann. Clin. Biochem.* **44**, 111–130 (2007).
15. Contributors to Wikimedia projects. Screening (medicine) - Wikipedia. *Wikimedia Foundation, Inc.* (2005). Available at: https://en.wikipedia.org/wiki/Screening_(medicine). (Accessed: 28th June 2019)
16. Akobeng, A. K. Understanding diagnostic tests 1: sensitivity, specificity and predictive values. *Acta Paediatrica* **96**, 338–341 (2007).
17. Eisenberg, M. J. Accuracy and predictive values in clinical decision-making. *Cleveland Clinic Journal of Medicine* **62**, 311–316 (1995).
18. Adams, M. C., GeneScreen Investigators, Evans, J. P., Henderson, G. E. & Berg, J. S. The promise and peril of genomic screening in the general population. *Genetics in Medicine* **18**, 593–599 (2016).
19. Green, R. C. *et al.* ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet. Med.* **15**, 565–574 (2013).
20. Burke, W. *et al.* Recommendations for returning genomic incidental findings? We need to talk! *Genet. Med.* **15**, 854–859 (2013).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

21. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).
22. Mammogram Accuracy - Accuracy of Mammograms. Available at: https://ww5.komen.org/BreastCancer/AccuracyofMammograms.html. (Accessed: 28th June 2019)
23. Rego, S. *et al.* High-frequency actionable pathogenic exome variants in an average-risk cohort. *Cold Spring Harb Mol Case Stud* **4**, (2018).
24. Jamuar, S. S. *et al.* Incidentalome from Genomic Sequencing: A Barrier to Personalized Medicine? *EBioMedicine* **5**, 211–216 (2016).
25. Olfson, E. *et al.* Identification of Medically Actionable Secondary Findings in the 1000 Genomes. *PLoS One* **10**, e0135193 (2015).
26. Dorschner, M. O. *et al.* Actionable, pathogenic incidental findings in 1,000 participants' exomes. *Am. J. Hum. Genet.* **93**, 631–640 (2013).
27. Manickam, K. *et al.* Exome Sequencing–Based Screening for BRCA1/2 Expected Pathogenic Variants Among Adult Biobank Participants. *JAMA Netw Open* **1**, e182140–e182140 (2018).
28. Abul-Husn, N. S. *et al.* Genetic identification of familial hypercholesterolemia within a single U.S. health care system. *Science* **354**, (2016).
29. OMIM Entry - # 145600 - MALIGNANT HYPERTHERMIA, SUSCEPTIBILITY TO, 1; MHS1. Available at: https://omim.org/entry/145600. (Accessed: 28th June 2019)
30. OMIM Entry - # 180200 - RETINOBLASTOMA; RB1. Available at: https://omim.org/entry/180200. (Accessed: 28th June 2019)
31. OMIM Entry - # 193300 - VON HIPPEL-LINDAU SYNDROME; VHL. Available at: https://omim.org/entry/193300. (Accessed: 28th June 2019)
32. OMIM Entry - # 158350 - COWDEN SYNDROME 1; CWS1. Available at: https://omim.org/entry/158350. (Accessed: 28th June 2019)
33. Tavtigian, S. V. *et al.* Modeling the ACMG/AMP variant classification guidelines as a

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

475 Bayesian classification framework. *Genetics in Medicine* **20**, 1054–1060 (2018).
476 34. Wikipedia contributors. Base rate fallacy. *Wikipedia, The Free Encyclopedia* (2019).
477 Available at:
478 https://en.wikipedia.org/w/index.php?title=Base_rate_fallacy&oldid=901526725.
479 (Accessed: 28th June 2019)
480 35. Ndase, P. *et al.* Frequency of false positive rapid HIV serologic tests in African men and
481 women receiving PrEP for HIV prevention: implications for programmatic roll-out of
482 biomedical interventions. *PLoS One* **10**, e0123005 (2015).
483 36. Diagnostic Test Calculator. Available at: http://araw.mede.uic.edu/cgi-bin/testcalc.pl.
484 (Accessed: 18th September 2019)
485 37. Diagnostic Test Calculator. *Diagnostic Test Calculator* Available at:
486 http://araw.mede.uic.edu/cgi-bin/testcalc.pl. (Accessed: 28th June 2019)
487 38. Mersch, J. *et al.* Prevalence of Variant Reclassification Following Hereditary Cancer
488 Genetic Testing. *JAMA* **320**, 1266–1274 (2018).
489 39. Macklin, S., Durand, N., Atwal, P. & Hines, S. Observed frequency and challenges of
490 variant reclassification in a hereditary cancer clinic. *Genet. Med.* **20**, 346–350 (2018).
491 40. Mayo Clinic GeneGuide - Results. (2018). Available at:
492 https://www.mayoclinic.org/mayoclinic-geneguide/results. (Accessed: 28th June 2019)
493 41. Murray, M. F. Your DNA is not your diagnosis: getting diagnoses right following
494 secondary genomic findings. *Genet. Med.* **18**, 765–767 (2016).
495 42. Media, H. Fortunately, Angelina Jolie Isn't On Medicare. Available at:
496 https://www.healthleadersmedia.com/finance/fortunately-angelina-jolie-isnt-medicare.
497 (Accessed: 17th August 2019)
498 43. Akbari, M. R., Gojska, N. & Narod, S. A. Coming of age in Canada: a study of population-
499 based genetic testing for breast and ovarian cancer. *Curr. Oncol.* **24**, 282–283 (2017).
500 44. Gabai-Kapara, E. *et al.* Population-based screening for breast and ovarian cancer risk
501 due to BRCA1 and BRCA2. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 14205–14210 (2014).
502 45. Metcalfe, K. A. *et al.* The risk of breast cancer in BRCA1 and BRCA2 mutation carriers

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

without a first-degree relative with breast cancer. *Clin. Genet.* **93**, 1063–1068 (2018).

46. Romero, J., Mejia-Lopez, E., Manrique, C. & Lucariello, R. Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC/D): A Systematic Literature Review. *Clin. Med. Insights Cardiol.* **7**, 97–114 (2013).
47. Akioyamen, L. E. *et al.* Estimating the prevalence of heterozygous familial hypercholesterolaemia: a systematic review and meta-analysis. *BMJ Open* **7**, e016461 (2017).
48. Khera, A. V. *et al.* Diagnostic Yield and Clinical Utility of Sequencing Familial Hypercholesterolemia Genes in Patients With Severe Hypercholesterolemia. *J. Am. Coll. Cardiol.* **67**, 2578–2589 (2016).
49. Reiner, Ž. Management of patients with familial hypercholesterolaemia. *Nat. Rev. Cardiol.* **12**, 565–575 (2015).
50. Rao, G. Remembering the meanings of sensitivity, specificity, and predictive values. *J. Fam. Pract.* **53**, 53 (2004).
51. Romero-Brufau, S., Huddleston, J. M., Escobar, G. J. & Liebow, M. Why the C -statistic is not informative to evaluate early warning scores and what metrics to use. *Crit. Care* **19**, 1–6 (2015).
52. Tenny, S. & Hoffman, M. R. Prevalence. in *StatPearls* (StatPearls Publishing, 2019).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

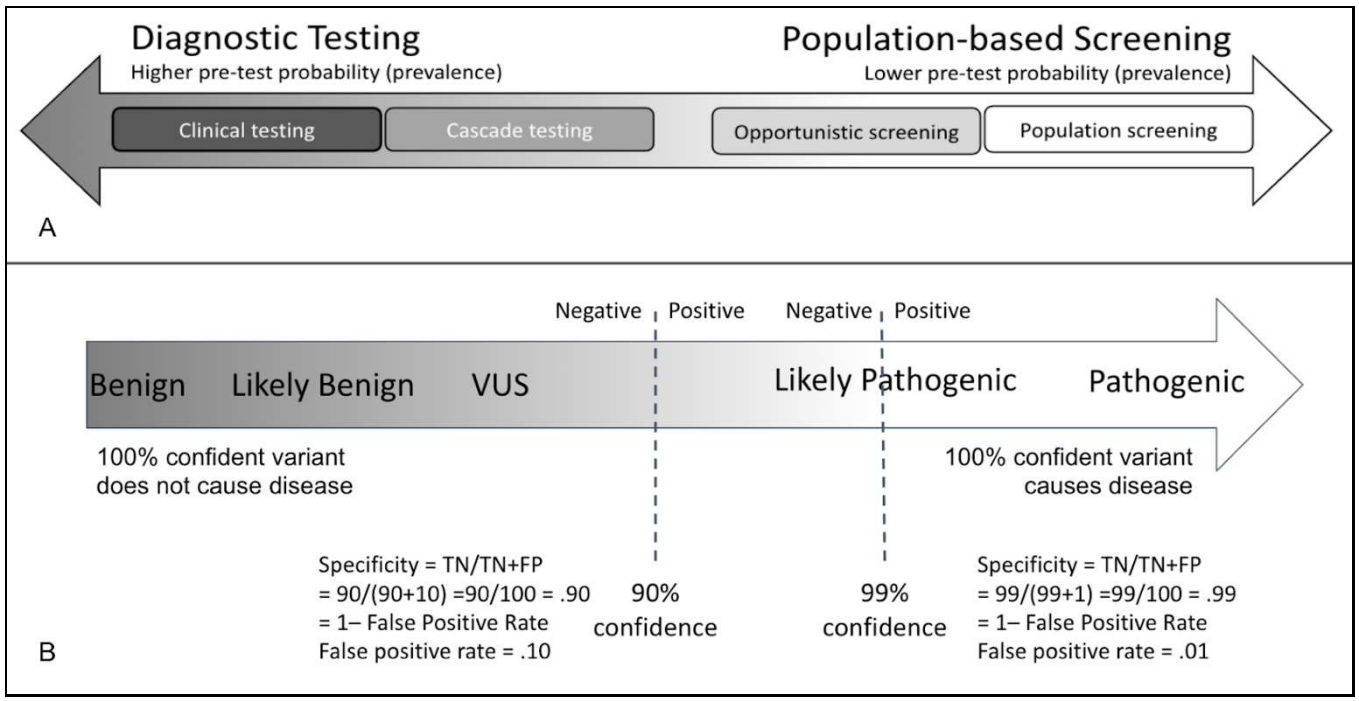


Figure 1. Intended Use and Cut-Off Optimization. A) The pre-test probability is equivalent to prevalence of the disease in the test population. Tests intended for use in high-risk populations will have a higher pre-test probability of disease than tests intended for screening unselected populations for unsuspected diseases. B) The clinical implications of false results must be considered when determining the appropriate sensitivity and specificity for an intended use. Tests intended for higher risk populations (diagnostic testing) may tolerate more false positives than tests intended for low-risk populations (screening).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

	Generic definitions	Genetics example
Analytical validity	The test accurately detects the analyte when it is present (analytical sensitivity) and does not detect it when it is absent (analytical specificity). Confirm by orthogonal technology during development phase.	The test accurately detects a sequence variant in *BRCA1* when it is present and does not detect it when it is absent. Confirm by orthogonal technology until performance metrics are well-established.
Clinical validity	The test accurately detects the disease when it is known to be present and does not detect it when it is known to be absent. A reference standard test is used to identify samples known to have the condition and those known to be free of the condition. During test development, these samples are the "truth set" against which the new test is compared in order to determine the test's clinical sensitivity and specificity.	The test accurately identifies samples from individuals known to have HBOC Syndrome and it does not identify HBOC Syndrome in samples from individuals known not to have HBOC Syndrome. A reference standard test is often not available to establish a truth set for use during development.
Clinical Sensitivity	The ability of the test to correctly identify those patients with the disease when present. (Note: This is different than diagnostic yield.)	"I know this person has HBOC. What is the chance that the test will show that this person has it?"[50]
Clinical Specificity	The ability of the test to correctly identify those patients without the disease when absent.	"I know this person doesn't have HBOC. What is the chance that the test will show that this person doesn't have it?"[50]
Positive Predictive Value	The probability that a person with a positive result has the condition. Dependent on the prevalence of the disease in the test population.	"I just received a positive HBOC test result for my patient. What is the chance that my patient actually has the disease?"[50] Dependent on the prevalence of the disease in the test population.
Negative Predictive Value	The probability that a person with a negative result does not have the condition. Dependent on the prevalence of the disease in the test population.	"I just received a negative HBOC test result for my patient. What is the chance that my patient actually doesn't have the disease?"[50] (Assume 100% analytical sensitivity) Dependent on the prevalence of the disease in the test population.
Penetrance	Not applicable.	The proportion of individuals with an inherited genetic syndrome (e.g., HBOC) who exhibit clinical symptoms (e.g., breast cancer) over time.
Classification	Evidence-based scoring system for determining likelihood of disease. Example Pap Smear: Cancer, Cervical Intraepithelial Neoplasia (3 levels), Atypical Squamous Cells of Uncertain Significance, Benign	Evidence-based scoring system for determining whether a variant is likely to cause disease. Example DNA Sequencing: Pathogenic, Likely pathogenic, Variant of uncertain significance, Likely Benign, Benign
Interpretation	Positive/Negative (Positives are actionable)	Positive/Negative (Positives are actionable)

Table 1. Generic test development definitions and examples of their application to genetic testing.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

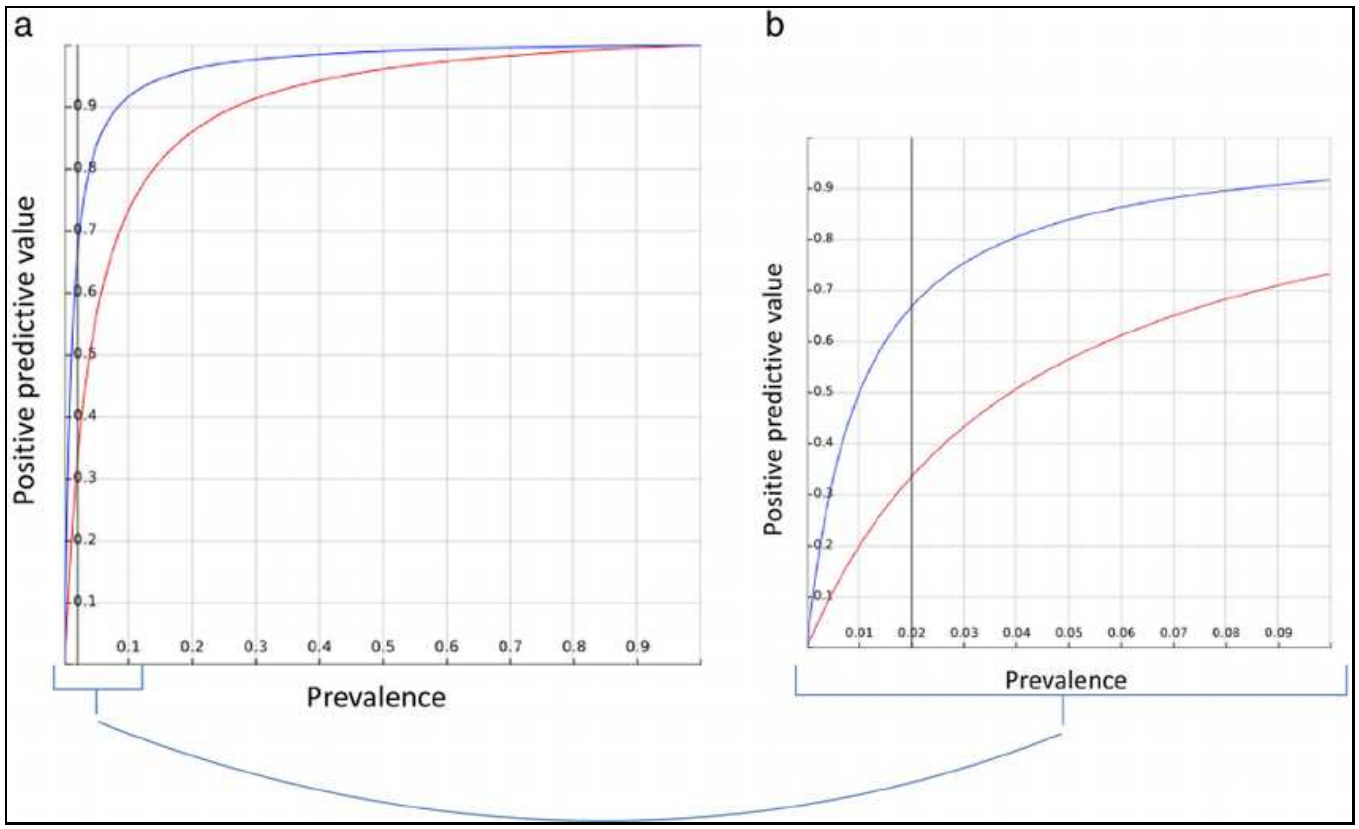


Figure 2. Impact of prevalence on PPV. When the disease prevalence is low in the test population, small changes in the specificity can have a large impact on the positive predictive value of the test. PPV as a function of prevalence for two tests: Test A (blue), with a sensitivity of 99% and a specificity of 99%; and Test B (red), with a sensitivity of 99% and a specificity of 96%. a) Full range of possible PPV and prevalence, from 0 to 1. b) Magnified region of prevalence <0.1, a gray line to show an example prevalence of 0.02. A decrease of only 3% in specificity can mean a 50% decrease in PPV: from 0.33 to 0.66. Adapted with permission from Romero-Brufau, et al. [51] The prevalence of the disease is equal to the *a priori* probability that a subject selected at random from the test population has the condition.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

Table 2: Estimating the Test Specificity					
Example	**Overall Positive rate**	**LP rate**	**Specificity**	**Prevalence**	**PPV Range**
CDC Tier 1	1.50%	0.50%	99.95%	1/200-1/500	80% - 91%
ACMG59™	3.00%	1.00%	99.90%	1/200-1/25,000	4% - 83%
ACMG59™ (at 6% overall positive rate)	6.00%	2.00%	99.79%	1/200-1/25,000	2% - 72%

Table 2. Estimating the Test Specificity. Test specificity is estimated by assuming ⅓ of the overall positive rate is due to likely pathogenic variants (LP) and then calculating the specificity as described in Supplemental Figure 2 (Specificity = TN/(TN+FP) = (1 - Positive Rate)/(1 - 29/30 x Positive Rate). Note how specificity changes with overall positive rate.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

Table 3: Computing the PPV (post-test probability) from prevalence and specificity							
	Disease Prevalence (pre-test probability)						
	1/25,000	**1/10,000**	**1/5000**	**1/1,000**	**1/500**	**1/200**	**1/50**
99.00%	0%	1%	2%	9%	17%	33%	67%
99.50%	1%	2%	4%	17%	29%	50%	80%
99.80%	*2%*	*5%*	*9%*	*33%*	*50%*	*72%*	91%
99.90%	*4%*	*9%*	*17%*	*50%*	*67%*	*83%*	95%
99.95%	7%	17%	29%	67%	80%	91%	98%
99.99%	44%	67%	80%	95%	98%	99%	100%

Table 3. Computing the PPV of likely pathogenic variants from prevalence and specificity.

The range of prevalences in the table are representative monogenic inherited diseases in an unselected population. The specificity for CDC Tier 1 conditions (underlined) and ACMG59™ (italicized) are from Table 2. The PPV calculations for CDC Tier 1 (underlined) and ACMG59™ (italicized) conditions are as follows: PPV = sensitivity x prevalence / [sensitivity x prevalence + (1 - specificity) x (1 - prevalence)].[52] The framework provided here is intended to provide estimates and show trends. In practice, each condition should be considered individually.

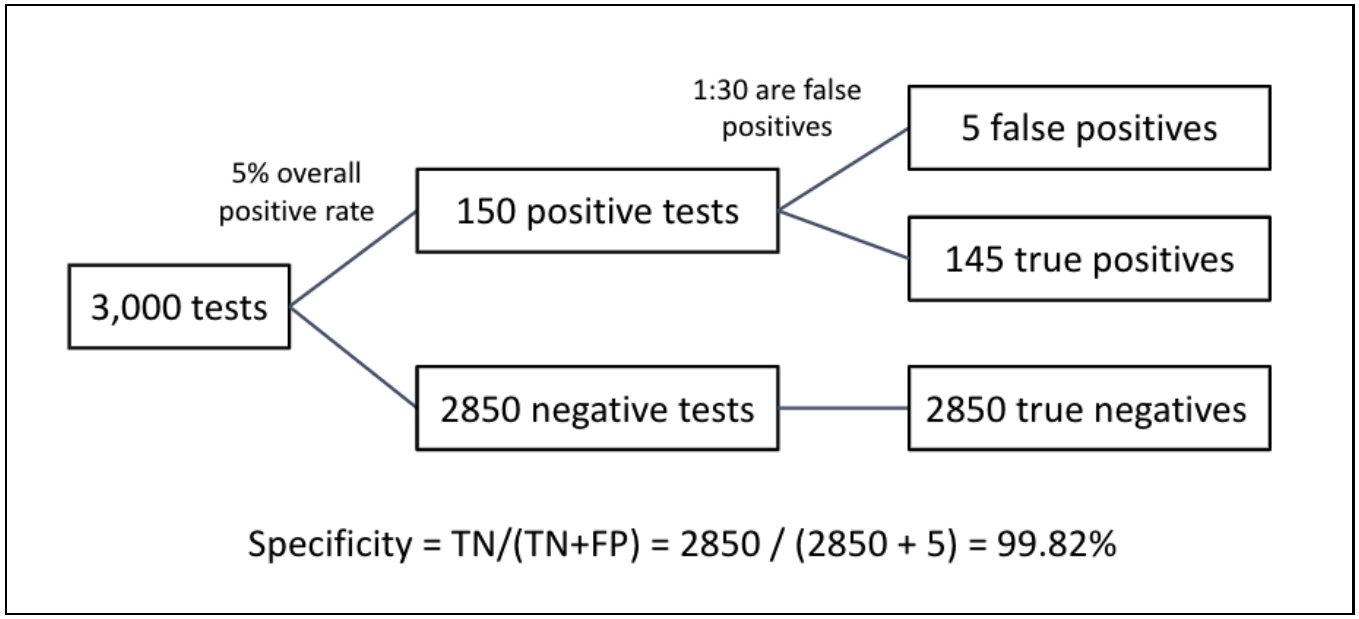
PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019

Supplemental Content

Screening test	Confirmatory test
Pap smear	Colposcopy
SickleDex or Thalassemia by genetic screen	Hemoglobin electrophoresis
HIV by ELISA	HIV by Western blot
Non-invasive prenatal testing (NIPT)	Karyotype or FISH
Genetic Factor XIII Deficiency or Glanzmann's	Mixing studies or Platelet Activation Test
Newborn screening by MS/MS	Biochemical testing, genetic testing

Supplemental Table 1. Screening test and confirmatory diagnostic test pairs. Examples of screening tests and confirmatory tests. Confirmatory tests can serve as the "gold standard" or reference standard test comparator during test validation as well as a follow up test for positive screening results in practice.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019



Supplemental Figure 1. Relationship between overall positive rate and specificity. In this example, the test has 100% analytical sensitivity and specificity, as determined during the test validation. After processing 3,000 samples, the lab determined that the positive rate in their setting is 5%. This means they have 150 positive results and 2850 negative results. The positive/negative cut off has been set to allow for 1/10 LP positives to be a false positive; thus 1/30 positives are false positives. Therefore, 5 of the positive results are false positives. and test specificity = 2850/(2850+5) = 99.82%. This method is used to calculate specificities for Tables 2 and 3 for varying positive test rates.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27922v3 | CC BY 4.0 Open Access | rec: 23 Oct 2019, publ: 23 Oct 2019